Polyhedral Newton-min algorithms for complementarity problems

Jean-Pierre Dussault, Mathieu Frappier, Jean Charles Gilbert

Polyhedral Newton-min algorithms for complementarity problems

Jean-Pierre Dussault [*], Mathieu Frappier [†], and Jean Charles Gilbert [‡]

Sunday 6^{th} October, 2019 (09:32)

The semismooth Newton method is a very efficient approach for computing a zero of a large class of nonsmooth equations. When the initial iterate is sufficiently close to a regular zero and the function is strongly semismooth, the generated sequence converges quadratically to that zero, while the iteration only requires to solve a linear system. If the first iterate is far away from a zero, however, it is difficult to force its convergence using linesearch or trust regions because a semismooth Newton direction may not be a descent direction of the associated least-square merit function, unlike when the function is differentiable. We explore this question in the particular case of a nonsmooth equation reformulation of the nonlinear complementarity problem, using the minimum function. We propose a globally convergent algorithm using a modification of a semismooth Newton direction that makes it a descent direction of the least-square function. Instead of requiring that the direction satisfies a linear system, it must be a feasible point of a convex polyhedron; hence, it can be computed in polynomial time. This polyhedron is defined by the often very few inequalities, obtained by linearizing pairs of functions that have close negative values at the current iterate; hence, somehow, the algorithm feels the proximity of a "bad kink" of the minimum function and acts accordingly. In order to avoid as often as possible the extra cost of having to find a feasible point of a polyhedron, a hybrid algorithm is also proposed, in which the Newton-min direction is accepted if a sufficient-descent-like criterion is satisfied, which is often the case in practice. Global convergence to regular points is proved; the notion of regularity is associated with the algorithm and is analysed with care.

Keywords: complementarity problem, global convergence, least-square merit function, linesearch, minimum function, nonsmooth reformulation, P-matrix, polyhedral Newton-min algorithm, semismooth Newton.

[*]Département d'Informatique, Faculté des Sciences, Université de Sherbrooke, Canada. E-mail: Jean-Pierre.Dussault@Usherbrooke.ca.

[†]Département de Mathématiques, Faculté des Sciences, Université de Sherbrooke, Canada. E-mail: Mathieu.Frappier@usherbrooke.ca.

[‡]INRIA Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France. E-mail: Jean-Charles.Gilbert@inria.fr.

Table of contents

1 Introduction

1.1 The complementarity problem

Let be given a positive integer n and two *smooth* functions $F : \Omega \to \mathbb{R}^n$ and $G : \Omega \to \mathbb{R}^n$ defined on an open subset Ω of $\mathbb{R}^n$. This paper considers, with an algorithmic point of view, the standard (nonlinear) complementarity problem. This problem consists in finding a vector $x \in \Omega$ such that

$$F(x) \geqslant 0, \qquad G(x) \geqslant 0, \qquad \text{and} \qquad F(x)^{\mathsf{T}} G(x) = 0, \tag{1.1a}$$

where vector inequalities must be taken in a componentwise fashion and $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto u^{\mathsf{T}} v = \sum_{i=1}^n u_i v_i$ is the Euclidean scalar product of $\mathbb{R}^n$ (the sign "$^{\mathsf{T}}$" is used to denote transposition of vectors and matrices). We denote by $[1:n] := \{1, \dots, n\}$ the set of the first n positive integers. In some contributions, the map G is supposed to be the identity; in addition to its generality, the model (1.1) presents the technical advantage of allowing us to avoid repeating reasonings, thanks to the possibility to switch F and G. Below, the system (1.1a) is written compactly as follows:

$$0 \leqslant F(x) \perp G(x) \geqslant 0, \tag{1.1b}$$

where the sign "$\perp$" refers to the required orthogonality of the vectors $F(x)$ and $G(x)$. The term "complementarity" comes from the fact that, due to the nonnegativity of $F(x)$ and $G(x)$ in (1.1), for all $i \in [1:n]$, either $F_i(x)$ or $G_i(x)$ must vanish and determining which of them is zero is part of the difficulty of the problem. The fact that these last conditions can be realized in 2^n different ways is at the origin of the complexity of the problem. It can be shown indeed that, even when the functions F and G are affine, finding a solution to (1.1) is NP-hard [20, 55; 1989-1991]. The algorithms considered in this paper can be easily adapted to the *mixed nonlinear complementarity problem*, in which the number p of complementarity conditions is less than the number n of unknowns, and there are $n-p$ additional nonlinear equality constraints. Less or more recent states of the art on the analysis of complementarity problems and numerical methods to solve then, in finite dimension, can be found in [66, 48, 70, 36, 24, 25, 50].

Occasionally, we shall make reference to the *linear complementarity problem* (LCP) in its standard form, which reads

$$0 \leqslant x \perp (Mx + q) \geqslant 0, \tag{1.2}$$

where the unknown is $x \in \mathbb{R}^n$, while $q \in \mathbb{R}^n$ and $M \in \mathbb{R}^{n \times n}$ are data. It corresponds to the nonlinear complementarity problem (1.1) with $F : x \mapsto Mx + q$ is affine and $G : x \mapsto x$ is the identity operator.

Complementarity conditions arise spontaneously in the first order optimality conditions of an optimization problem with inequality constraints and these conditions can be written like the system (1.1). The complementarity system (1.1) is also often used to model in part problems in which several systems of equations are, to some extend, in competition. The one that is active in a given place and at a given time, corresponding to a common index of $F(x)$ and $G(x)$, depends on threshold effects; if the threshold $F_i(x) = 0$ is not reached, i.e., $F_i(x) > 0$, then the equation $G_i(x) = 0$ is active, and vice versa. Examples include problems in nonsmooth mechanics and dynamics [1, 15], the phase transition problem in multiphase flows [64, 65, 11, 26, 17, 4, 6, 7], precipitation-dissolution problems in chemistry [16, 57], portfolio management in finance [43], computer graphics [35], meteorology simulation, economic equilibrium, to mention a few. Surveys on examples of applications of the complementarity problem can be found in [46, 48, 70, 38, 36].

1.2 A few linearization algorithms

Many techniques have been proposed to solve (1.1) since the problem was introduced by Cottle in his PhD thesis in 1964 [22, 23]. It is out of the scope of this paper to review all of them and we refer instead the interested reader to the recent monographs [36, 50]. Below, we limit our account to the algorithms in close connection with the numerical methods proposed and analyzed in this paper. The motivation is to put in perspective the proposed algorithms, essentially within the Newton-min-type methods. On the way, we introduce notation and concepts used throughout the paper.

The adjacent numerical methods are related to the Newton algorithm to solve the nonsmooth system of equations

$$H(x) = 0, \tag{1.3a}$$

in which $H : \Omega \to \mathbb{R}^n$ is the function defined at $x \in \Omega$ by

$$H(x) := \min(F(x), G(x)), \tag{1.3b}$$

where the minimum is taken componentwise [2, 69]. It is clear that problems (1.1) and (1.3) have the same solutions, since, for two real numbers a and b, $\min(a, b) = 0$ if and only if $a \geqslant 0$, $b \geqslant 0$, and $ab = 0$ (for other functions having that property, see [62, 42] and the references therein). The term "Newton-min" was coined in [8, 9, 10] to name this solution strategy and we adopt it in this paper. The proposed methods are globalized by using the classical merit function associated with H [67, 31, 13, 14], which is the least-square function $\theta : \Omega \to \mathbb{R}$ defined at $x \in \Omega$ by

$$\theta(x) := \frac{1}{2}\,\|H(x)\|^2 = \frac{1}{2}\,\|\min(F(x), G(x))\|^2. \tag{1.4}$$

The goal of this paper is to focus on the reformulation (1.3) and its globalization, using linesearch on the natural merit function (1.4). More is said on the proposed approaches in section 1.3 below, after the presentation of some related linearization methods.

Many other equation reformulations of the complementarity problem have been proposed, see [62, 27, 53, 52, 77, 18, 28, 39, 37, 71, 47] and the references therein. Our choice of a reformulation by the minimum function is not only motivated by an intellectual curiosity (as we shall see, there are still holes in its implementation and its analysis), but also by its observed efficiency. This one is sometimes explained by the piecewise affine nature of the minimum function, which provides no additional nonlinearity besides its nondifferentiability. From a theoretical point of view, the required regularity of the solution to guarantee fast local convergence of a Newton-like algorithm on (1.3) is also less restrictive than on the Fischer-Burmeister reformulation, for instance, and this algorithm has finite termination for the linear complementarity problem, which cannot be expected when the reformulation is more nonlinear [36; § 9.2].

A first linearization method to solve (1.1) consists in applying Josephy-Newton (JN) iterations [51] on a functional inclusion reformulation of the problem [54] (see [36; § 7.3] for a reformulation using the normal map). This results in linearizing the functions in (1.1b) while keeping its complementarity problem structure: the new iterate $x + d$, following the current one x, is determined by taking for d an appropriate solution to the linear complementarity problem in d (if this solution exists)

$$0 \leqslant \big(F(x) + F'(x)d\big) \perp \big(G(x) + G'(x)d\big) \geqslant 0. \tag{1.5}$$

The SQP algorithm in nonlinear optimization can also be derived by this technique [51], so that the two methods have common features. The local quadratic convergence of this algorithm can be deduced from the one of the JN iterations for a functional inclusion (Josephy [51] assumes that the sought solution is strongly regular in the sense of Robinson [74], while Bonnans [12] only assumes the weaker so-called semistability and hemistability; see also [36; § 7.3] for related results). The globalization of this linearization approach for complementarity problems uses adapted merit functions (see [61] and the references therein). The JN approach has many attractive features, but, with respect to the methods proposed in section 1.3, the system (1.5) has the inconvenient of requiring the computation of a solution to a linear complementarity problem of dimension n at each iteration and we have already mentioned that such a problem is generally NP-hard. We also point out that this approach is not relevant in the case when the original problem (1.1) is a linear complementarity problem, since then (1.5) is exactly the same problem as the original one.

Another linearization approach to solve (1.1) consists in applying a Newton-like method to solve directly the equivalent nonsmooth system (1.3). Among these methods, one finds the B-Newton algorithm [68], which is adapted to B-differentiable maps [30, 75, 76]. For a locally Lipschitz function defined on a space of finite dimension, like H in (1.3b), the B-derivative is identical to the directional derivative [75, 76], so that the direction d giving the new iterate $x + d$ in the B-Newton algorithm is taken as a solution (if any) to the (usually nonlinear) system

$$H(x) + H'(x; d) = 0, \tag{1.6}$$

where $H'(x; d) := \lim_{t\downarrow 0}[H(x + td) - H(x)]/t$ is the usual one-side directional derivative. It is plain to see that the function H given by (1.3b) is directionally differentiable (recall that F and G are supposed to be smooth) and that its directional derivative is given by

$$H_i'(x; d) = \begin{cases} F_i'(x)d & \text{if } i \in \mathcal{F}(x), \\ G_i'(x)d & \text{if } i \in \mathcal{G}(x), \\ \min(F_i'(x)d, G_i'(x)d) & \text{if } i \in \mathcal{E}(x), \end{cases} \tag{1.7}$$

where we have used the following mnemonic notation for index sets, which will be often useful below:

$$\begin{array}{rcl} \mathcal{F}(x) & := & \{i \in [1:n] : F_i(x) < G_i(x)\}, \\ \mathcal{G}(x) & := & \{i \in [1:n] : F_i(x) > G_i(x)\}, \\ \mathcal{E}(x) & := & \{i \in [1:n] : F_i(x) = G_i(x)\}. \end{array} \tag{1.8}$$

Combining (1.6), (1.3b), and (1.7), we see that the search direction d of the *B-Newton-min algorithm* is determined as a solution (if any) to the system

$$\begin{cases} (F(x) + F'(x)d)_{\mathcal{F}(x)} = 0, \\ (G(x) + G'(x)d)_{\mathcal{G}(x)} = 0, \\ 0 \leqslant (F(x) + F'(x)d)_{\mathcal{E}(x)} \perp (G(x) + G'(x)d)_{\mathcal{E}(x)} \geqslant 0. \end{cases} \tag{1.9}$$

An interesting asset of the B-Newton-min approach, compared to the JN algorithm, is that the system (1.9) can be much easier to solve than (1.5), since its number $|\mathcal{E}(x)|$ of complementarity conditions is reduced to the number of indices i giving the equality $F_i(x) = G_i(x)$ at the current x and that this number can be very small. The convergence properties of this algorithm based on (1.9) derive from the one of the B-Newton algorithm (1.6) for solving the equation $H(x) = 0$, with a B-differentiable function H. According to [68; theorem 3], the algorithm converges when the first iterate is in some neighborhood of a zero x_* of H at which H is strongly Fréchet differentiable with a nonsingular $H'(x_*)$; this required smoothness assumption on H is awkward and rather restrictive when one aims at solving a nonsmooth system. Another interesting asset of the B-Newton direction d is that it is a descent direction of θ at x [68; lemma 1], which gives rise to a linesearch algorithm, generating sequences whose accumulation points x_* are solutions to (1.3a), provided H is strongly Fréchet differentiable at x_* and $H'(x_*)$ is injective [68; theorem 4(iii)]; these are again rather restrictive assumptions. In terms of the data of problem (1.1), when G is the identity, these conditions are guaranteed if the accumulation point x_* is regular in the sense of [68; definition 2] and $(x_*)_i = F_i(x_*) = 0$ for $i \in \mathcal{E}(x_*)$ [68; theorem 6]. Finally, we point out that the B-Newton-min is not appropriate to solve the linear complementarity problem (1.2), since (1.9) is identical to the original problem when $\mathcal{E}(x) = [1\,{:}\,n]$.

The B-Newton-min algorithm is modified in [69] in order to obtain convergence results with less demanding assumptions and the modification is shown in [45] to be part of a larger family of globally convergent algorithms for solving a nonsmooth system $H(x) = 0$. In the case of problem (1.1), the modified B-Newton-min algorithm consists in computing the new iterate $x + d$, from the current one x, by determining d as a solution (if any) to the nonlinear system [45; (4)]

$$H(x) + D(x, d) = 0, \tag{1.10}$$

where $D : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is no longer the directional derivative of H like in (1.6)-(1.7) but is defined by [45; (12)]

$$D_i(x, d) = \begin{cases} F_i'(x)d & \text{if } F_i(x) < G_i(x),\ G_i(x) \geqslant 0, \\ G_i'(x)d & \text{if } F_i(x) > G_i(x),\ F_i(x) \geqslant 0, \\ \min(F_i'(x)d, G_i'(x)d) & \text{otherwise.} \end{cases} \tag{1.11}$$

In comparison with (1.6), we see that some indices of $\mathcal{F}(x)$ and $\mathcal{G}(x)$ are now treated like those of $\mathcal{E}(x)$. Rewriting (1.10), with the form of H from (1.3b) and that of D from (1.11), we see that d has to solve the system

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } F_i(x) < G_i(x),\ G_i(x) \geqslant 0, \\ G_i(x) + G_i'(x)d = 0 & \text{if } F_i(x) > G_i(x),\ F_i(x) \geqslant 0, \\ 0 \leqslant (F_i(x) + F_i'(x)d) \perp (F_i(x) + G_i'(x)d) \geqslant 0 & \text{if } F_i(x) < G_i(x) < 0, \\ 0 \leqslant (G_i(x) + F_i'(x)d) \perp (G_i(x) + G_i'(x)d) \geqslant 0 & \text{if } 0 > F_i(x) > G_i(x), \\ 0 \leqslant (F_i(x) + F_i'(x)d) \perp (G_i(x) + G_i'(x)d) \geqslant 0 & \text{otherwise.} \end{cases} \tag{1.12}$$

This eclectic system has therefore more complementarity conditions than (1.9), but has also better convergence results. Conditions ensuring the existence and uniqueness of the solution to the mixed linear complementarity problem (1.12) can be obtained [69; § 5]. Furthermore, it can be shown that this direction d is a descent direction of θ at x, which gives rise to a linesearch algorithm whose global convergence (without the previously required smoothness of H) and the admissibility of the unit stepsize are studied in [69; §§ 6-8]. For the same reason as for the B-Newton-min algorithm, the present modification is not appropriate for linear complementarity problem (LCP), since for some such problems, the system (1.12) may be identical to the original problem.

A more drastic approach to solve a nonsmooth system $H(x) = 0$ is to use the semismooth Newton method [73, 72], provided H is semismooth. This method only requires to solve a linear system per iteration: one chooses a Jacobian J_x in the generalized Clarke's differential $\partial_C H(x)$ of H at x [21] and defines the displacement d at x as a solution (if any) to

$$H(x) + J_x d = 0. \tag{1.13}$$

Despite its poor description of the function H at a point of nondifferentiability, this method has the remarkable property of having a superlinear speed of convergence (or quadratic, if H is strongly semismooth), when the first iterate is close enough to a *regular point* x_* of H, which means here that all the Jacobians of $\partial_C H(x_*)$ are nonsingular [36, 50]. A drawback of this method is that it is often difficult to compute an element of $\partial_C H(x)$, for a particular function H, because this generalized Jacobian is not known or evaluating one of its elements is computationally expensive. Nevertheless, one can sometimes use a surrogate of the generalized Jacobian J_x in (1.13), while keeping the fast local convergence property of the pure approach (see [44, 59] for the projection on a convex polyhedron). A method inspired from the semismooth Newton algorithm or from [56], applied to (1.3), computes the displacement d from x to the next iterate $x + d$ by solving (if possible) the linear system

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{F}}(x), \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{G}}(x), \end{cases} \tag{1.14}$$

where the pair $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ forms a partition of $[1:n]$ and satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$. The drawback of this economical approach, however, is that d is not necessary a descent direction of the natural least-square merit function θ, because of an inappropriate choice of the indices of $\mathcal{E}(x)$ going into $\tilde{\mathcal{F}}(x)$ and $\tilde{\mathcal{G}}(x)$ (see counter-example 2.4 below, for a linear complementarity problem), which explains why there is no globally convergent algorithm based on this direction and the merit function (1.4) (there are hybrid approaches, however, using the merit functions (1.4) and the Fischer-Burmeister merit function [60, 28, 71]).

1.3 A foretaste of the proposed algorithms

The methods proposed and analyzed in this paper are progressively introduced in section 2, but we can already give here a foretaste of their nature. They find their place in the panorama of linearization methods of the minimum function (1.3b) presented in the previous section, in the sense that their directions can be viewed as intermediates between the B-Newton direction d given by (1.9), or its modification given by (1.12), and the semismooth-like direction computed by (1.14), called the *plain Newton-min* direction in section 2.1. Their main advantage is to avoid the need of solving an LCP at each iteration, hence unlike in (1.9) or (1.12), and to guarantee global convergence, hence unlike (1.14).

Instead of having to solve an LCP, the direction must satisfy a system, made of affine equalities and (generally very few) *inequalities*, in order to guarantee the descent of the least-square merit function θ, defined in (1.4); see section 2.2. A least-norm displacement of this system can, for example, be obtained by solving a convex quadratic optimization problem, which can be done in polynomial time. An improvement of this direction is needed, however, to guarantee convergence in the sense and with the technique of proof presented in section 3.2: the set of inequalities defining the direction must be slightly enlarged when the iterate is near a "bad kink" of H; see section 2.3. Finally, to avoid the more expensive direction, due to the presence of inequalities in its definition, a hybrid algorithm is proposed in section 2.4, in which the descent property of the semismooth direction (1.14) is first tested: if a sufficient decrease along that direction is ensured, this one is adopted by the algorithm.

Like any linearization algorithm with linesearch, convergence is restricted by a regularity assumption on the limit point. This notion of regularity depends on the computed direction. This issue is analyzed with care in section 3.1. Finally, a global convergence result is given in section 3.2. The paper ends with the conclusion section 4.

The design of the algorithms presented in this paper has been oriented by an intensive numerical exploration, which has shown that the proposed method is competitive with other solvers on various applications, on some reference academic examples, and on randomly generated problems. These experiments are reported in [40, 33] for the linear complementarity problem (1.2).

1.4 Notation

We denote by $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^n$. The cardinality of a set S (i.e., its number of elements, which will be always finite) is denoted by $|S|$.

2 Polyhedral Newton-min directions

This section introduces the directions of the proposed algorithms. It proceeds gradually, insisting on the motivation, which is to obtain descent directions of θ and to guarantee some global convergence property. We first observe that the plain Newton-min (NM) direction of section 2.1, already presented in (1.14) and obtained by solving a single linear system, is not necessarily a descent direction of θ (counter-example 2.4). We then examine in section 2.2 the reason of this descent property failure and propose a descent direction (proposition 2.5), which must satisfy a similar system as the one of the plain NM direction, but whose equations corresponding to the indices in $\{i \in [1:n] : F_i(x) = G_i(x) < 0\}$ are transformed into pairs of inequalities. This yields what we call a *polyhedral Newton-min* (PNM) direction since this one must be a feasible point of a certain polyhedron. This *plain PNM direction* is always a descent direction of θ. Nevertheless, it did not allowed us to prove the global convergence result of theorem 3.18 for a reason discussed in section 2.3. It seems important, indeed, that, when the current iterate is near some kinks of H (not all of them), the direction is built by picking information on the behavior of the function H on both sides of the kink. This leads us to propose in section 2.3.1 the *secured* PNM direction (2.12), whose definition depends on the proximity of the current iterate to these special kinks of H. Its descent property is viewed in section 2.3.2 as a consequence of proposition 2.7, which analyses the potential descent property of a direction by averaging its effect on each term $H_i(x)^2$ defining the merit function θ. Section 2.3.2 also introduces the very permissive *inexact secured* PNM direction (2.22), for which descent property and global convergence hold, but that is too expensive to compute in the present context. We conclude with section 2.4, which presents the *hybrid Newton-min direction* and the assoiated *hybrid Newton-min algorithm.* This algorithm takes the plain NM direction (because it is cheap to compute) if this one can ensure a sufficient decrease of the merit function θ (this is not guaranteed) or, otherwise, it computes a more expensive secured PNM direction. Both the secured PNM algorithm and the hybrid PNM algorithm have their global convergence analyzed in section 3.2.

2.1 Plain Newton-min directions

The *plain Newton-min* (NM) *algorithm* is a semismooth Newton-like method on the reformulation (1.3) of the nonlinear complementarity problem (1.1), which uses the minimum function (algorithm 7.2.17 in [36]). It computes its direction d at $x \in \Omega$ by solving the linear system (1.14), which is reproduced here for the reader's convenience:

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{F}}(x), \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{G}}(x). \end{cases} \tag{2.1}$$

In this system, the pair $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ forms a partition of $[1:n]$ and satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$. By the "symmetry" in F and G of the complementarity problem (1.1), there is no natural reason to put all the indices of $\mathcal{E}(x)$ in $\tilde{\mathcal{F}}(x)$ or $\tilde{\mathcal{G}}(x)$, which motivates the flexibility admitted in the direction definition (2.1). We see that, at a point x on a possible kink of H, due to one of its components $i \in \mathcal{E}(x)$, a pseudo-Jacobian of H_i at x is chosen in $\{F_i'(x), G_i'(x)\}$.

To identify the points x at which the linear system (2.1) is guaranteed to have a solution, we introduce the notion of *NM-regularity.* This notion is linked to the plain NM algorithm, like the nonsingularity of the Jacobian of a nonlinear system is a regularity assumption linked to Newton's method.

Definition 2.1 (NM-regularity) A point $x \in \mathbb{R}^n$ is said to be *NM-regular* (we also say that the complementarity problem (1.1) is *NM-regular* at $x \in \mathbb{R}^n$) if F and G are differentiable at x and if, for any partition $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ of $[1:n]$ that satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$, the Jacobian

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}(x)}(x) \\ G'_{\tilde{\mathcal{G}}(x)}(x) \end{pmatrix} \tag{2.2}$$

is nonsingular. □

When G is the identity, one recovers the notion of b-regularity of [41; definition 2].

The NM-regularity of a point diffuses to the neighboring points.

Proposition 2.2 (diffusing NM-regularity) *Suppose that F and G are differentiable near some $\bar{x} \in \mathbb{R}^n$, that F' and G' are continuous at $\bar{x}$, and that $\bar{x}$ is NM-regular. Then, any x near $\bar{x}$ is NM-regular.*

Proof. By their differentiability property, F and G are continuous at $\bar{x}$. Then, it immediately follows that, for x near $\bar{x}$:

$$\mathcal{F}(\bar{x}) \subseteq \mathcal{F}(x) \qquad \text{and} \qquad \mathcal{G}(\bar{x}) \subseteq \mathcal{G}(x).$$

Suppose now that $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ is a partition of $[1\!:\!n]$ that satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$. By the preceding inclusions, $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ is a partition of $[1\!:\!n]$ that satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(\bar{x})$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(\bar{x})$. The NM-regularity at $\bar{x}$ now implies that the matrix

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}(x)}(\bar{x}) \\ G'_{\tilde{\mathcal{G}}(x)}(\bar{x}) \end{pmatrix}$$

is nonsingular. Since the set of nonsingular linear operators is open and since F' and G' are continuous at $\bar{x}$, it follows that

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}(x)}(x) \\ G'_{\tilde{\mathcal{G}}(x)}(x) \end{pmatrix} \tag{2.3}$$

is nonsingular for x near $\bar{x}$. As a result, neighboring x's are NM-regular. □

The next property will be useful for establishing the global convergence result of theorem 3.18 (see [41; lemma 3] for a similar property).

Proposition 2.3 (local boundedness of NM directions) *Suppose that F and G are differentiable near some $\bar{x} \in \mathbb{R}^n$, that F' and G' are continuous at $\bar{x}$, and that $\bar{x}$ is NM-regular. Then, there is a constant C and a neighborhood V of $\bar{x}$, such that, for all $x \in V$, the system (2.1) has a unique solution d bounded normwise by C.*

Proof. By proposition 2.2, if x is near $\bar{x}$ and if $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ is a partition of $[1\!:\!n]$ that satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$, then the operator in (2.3) is nonsingular. By restricting the neighborhood V of $\bar{x}$ on which this property is verified, one can also guarantee that the operators in (2.3), for $x \in V$, have bounded inverses (Banach perturbation lemma). Then, the directions d uniquely defined by (2.1) are also bounded on a possibly smaller neighborhood of $\bar{x}$ (to get the boundedness of $F(x)$ and $G(x)$). □

The plain NM direction is very attractive since it can be computed by solving a single linear system and because it guarantees a local quadratic convergence [56, 58]. Unfortunately, this direction may not be a descent direction of the least-square merit function θ defined in (1.4), although this one is naturally associated with the system (1.3). Here is an example of this phenomenon in the case of a linear complementarity problem with a **P**-matrix (this fact was already observed during the preparation of the PhD thesis of I. Ben Gharbia [5; example 5.8]).

Counterexample 2.4 (no descent direction from (2.1)) Consider the linear complementarity problem (1.2) in dimension $n = 2$ and the point x given by

$$M = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}, \qquad q = \begin{pmatrix} -4 \\ -2 \end{pmatrix}, \qquad \text{and} \qquad x = \begin{pmatrix} -2 \\ 1 \end{pmatrix}. \tag{2.4}$$

Since $F(x) \equiv x = (-2, 1)$ and $G(x) \equiv Mx + q = (-2, -1)$, the index sets (1.8) read $\mathcal{F}(x) = \varnothing$, $\mathcal{G} = \{2\}$, and $\mathcal{E}(x) = \{1\}$. If one computes the NM direction d by (2.1) with $\tilde{\mathcal{F}} = \varnothing$ and $\tilde{\mathcal{G}} = \{1, 2\}$, one gets $d = -x - M^{-1}q = (-2, 1)$. Then, for $t \geqslant 0$:

$$\theta(x + td) = \frac{5t^2 + 6t + 5}{2} \qquad \text{and} \qquad \theta'(x; d) = 3,$$

which shows that the chosen NM direction d is an ascent direction of θ at x. The increase of θ along the chosen NM direction is due to a wrong choice of $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$. By choosing the index sets $\tilde{\mathcal{F}} = \{1\}$ and $\tilde{\mathcal{G}} = \{2\}$, one gets the solution $d = (2, 1)$ to the linear complementarity problem (1.9) and $x + d = (0, 2)$ is the solution to the LCP. □

To the best of our knowledge, this intrinsic difficulty of the plain NM algorithm has not been considered with full attention (we quote, however, algorithm 9.2.2 in [36], which requires to solve a convex quadratic optimization problem at each iteration with n bound constraints and is therefore more expensive than the algorithms proposed below). In sections 2.2 and 2.3, we propose to overcome the difficulty by imposing the direction to be a feasible point of particular polyhedrons, with a very small number of inequality constraints, instead of being the solution to a linear system. The computation of the direction is therefore more expensive, but remains polynomial. In addition, in sections 2.4, a heuristics is proposed to avoid as much as possible the need to find a point in this polyhedron.

2.2 Plain polyhedral Newton-min directions

The direction proposed in this section is based on the following computation, which highlights the reason why a plain NM direction may not be a descent direction of the least-square merit function θ defined in (1.4). First, observe that the map θ is directionally differentiable as a composition of H, which is directionally differentiable, and $\frac{1}{2}\|\cdot\|^2$ which is Lipschitz continuous and smooth. In this case, the chain rule applies (see [14; lemma 11.1] for example):

$$\theta'(x; d) = H(x)^{\mathsf{T}} H'(x; d). \tag{2.5}$$

Since near x, $H_{\mathcal{F}} \equiv F_{\mathcal{F}}$ and $H_{\mathcal{G}} \equiv G_{\mathcal{G}}$, it is natural to impose to a Newton-like direction d to verify

$$(F(x) + F'(x)d)_{\mathcal{F}(x)} = 0 \qquad \text{and} \qquad (G(x) + G'(x)d)_{\mathcal{G}(x)} = 0. \tag{2.6}$$

Note, however, that it will be necessary to infringe this rule below, in order to approach some kinks of H with caution. Now, from (2.5), (1.3b), and (1.7), the directional derivative of θ reads

$$\begin{aligned}\theta'(x; d) \;&=\; F_{\mathcal{F}(x)}(x)^{\mathsf{T}} F'_{\mathcal{F}(x)}(x)d + G_{\mathcal{G}(x)}(x)^{\mathsf{T}} G'_{\mathcal{G}(x)}(x)d \\ &\quad\; + F_{\mathcal{E}(x)}(x)^{\mathsf{T}} \min(F'_{\mathcal{E}(x)}(x)d, G'_{\mathcal{E}(x)}(x)d).\end{aligned}$$

Next, using (2.6) and $F_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x)$, the directional derivative $\theta'(x; d)$ becomes

$$\begin{aligned}\theta'(x; d) \;&=\; -\|F_{\mathcal{F}(x)}(x)\|^2 - \|G_{\mathcal{G}(x)}(x)\|^2 - \|F_{\mathcal{E}(x)}(x)\|^2 \\ &\quad\; + F_{\mathcal{E}(x)}(x)^{\mathsf{T}} \min(F_{\mathcal{E}(x)}(x) + F'_{\mathcal{E}(x)}(x)d, G_{\mathcal{E}(x)}(x) + G'_{\mathcal{E}(x)}(x)d) \\ &=\; -2\,\theta(x) + F_{\mathcal{E}(x)}(x)^{\mathsf{T}} \min(F_{\mathcal{E}(x)}(x) + F'_{\mathcal{E}(x)}(x)d, G_{\mathcal{E}(x)}(x) + G'_{\mathcal{E}(x)}(x)d).\end{aligned}$$

The first term in the right-hand side is satisfactory since it corresponds to the formula of the directional derivative in the smooth case, while the second term is at the origin of the positive directional derivative observed in counter-example 2.4. Let us dissect this last term in order to see what conditions the direction must verify to make it nonpositive (we take up again an observation already made during the preparation of the PhD thesis of I. Ben Gharbia [5; 2012] for the LCP (1.2)). For this, we introduce the following partition $(\mathcal{E}^-(x), \mathcal{E}^0(x), \mathcal{E}^+(x))$ of $\mathcal{E}(x)$:

$$\begin{array}{rcl} \mathcal{E}^-(x) & := & \{i \in [1:n] : F_i(x) = G_i(x) < 0\}, \\ \mathcal{E}^0(x) & := & \{i \in [1:n] : F_i(x) = G_i(x) = 0\}, \\ \mathcal{E}^+(x) & := & \{i \in [1:n] : F_i(x) = G_i(x) > 0\}, \\ \mathcal{E}^{0+}(x) & := & \{i \in [1:n] : F_i(x) = G_i(x) \geqslant 0\}, \\ \mathcal{E}^{0-}(x) & := & \{i \in [1:n] : F_i(x) = G_i(x) \leqslant 0\}. \end{array} \tag{2.7}$$

Let $i \in \mathcal{E}(x) = \mathcal{E}^{0+}(x) \cup \mathcal{E}^-(x)$.

- If $i \in \mathcal{E}^{0+}(x)$, then $F_i(x) \geqslant 0$. If one of the linearized functions $F_i(x) + F_i'(x)d$ or $G_i(x) + G_i'(x)d$ vanishes, their minimum is nonpositive, yielding $F_i(x)\min(F_i(x) + F_i'(x)d, G_i(x) + G_i'(x)d) \leqslant 0$.
- If $i \in \mathcal{E}^-(x)$, then $F_i(x) < 0$. To get $F_i(x)\min(F_i(x) + F_i'(x)d, G_i(x) + G_i'(x)d) \leqslant 0$, it is now necessary to have $\min(F_i(x) + F_i'(x)d, G_i(x) + G_i'(x)d) \geqslant 0$ meaning that the following inequalities must hold:

$$F_i(x) + F_i'(x)d \geqslant 0 \quad \text{and} \quad G_i(x) + G_i'(x)d \geqslant 0. \tag{2.8}$$

Therefore, the decrease of θ is ensured along a direction d if this one satisfies (2.6), either $F_i(x) + F_i'(x)d = 0$ or $G_i(x) + G_i'(x)d = 0$ when $i \in \mathcal{E}^{0+}(x)$, and both inequalities in (2.8) for $i \in \mathcal{E}^-(x)$.

The above discussion leads us to the definition of the following direction. Let us denote by $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ an arbitrary partition of $\mathcal{E}^{0+}(x)$, meaning that

$$\mathcal{E}^{0+}(x) = \mathcal{E}^{0+}_{\mathcal{F}}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) \qquad \text{and} \qquad \mathcal{E}^{0+}_{\mathcal{F}}(x) \cap \mathcal{E}^{0+}_{\mathcal{G}}(x) = \varnothing. \tag{2.9}$$

A *plain polyhedral Newton-min* (PNM) *direction* is a direction d that satisfies the following system

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x) \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) \\ F_i(x) + F_i'(x)d \geqslant 0 & \text{if } i \in \mathcal{E}^-(x) \\ G_i(x) + G_i'(x)d \geqslant 0 & \text{if } i \in \mathcal{E}^-(x). \end{cases} \tag{2.10}$$

Note that $(\mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x), \mathcal{E}^-(x))$ forms a partition of $[1:n]$. Therefore, we have imposed inequality constraints on the linearized functions $F_i(x) + F_i'(x)d$ and $G_i(x) + G_i'(x)d$ for the indices in $i \in \mathcal{E}^-(x)$, like suggested by (2.8), instead of forcing one arbitrary of them to vanish, like in the plain NM algorithm (2.1).

The computation of a plain PNM direction is more expensive than the computation of the plain NM direction (2.1), since a feasible point of a convex polyhedron must be found instead of the solution to a linear system. Nevertheless, a direction satisfying (2.10) can be computed in polynomial time using linear or quadratic optimization (see [29, 14, 19] and the references therein). Such an extra cost is acceptable, even when one solves a linear complementarity problem. In the next section, we continue to explore this vein and in section 2.4, we introduce a way to reduce the cost of the direction computation which is very successful in practice.

We summarize the discussion of this section in the following proposition.

Proposition 2.5 (descent property with (2.10)) *For any direction d satisfying (2.10), one has $\theta'(x;d) \leqslant -2\,\theta(x)$.*

2.3 Secured polyhedral Newton-min directions

Although a vector d satisfying (2.10) is a descent direction of θ, we were not able to get the global convergence result of theorems 3.17 and 3.18 below with that direction. In the approach followed in the proof of theorem 3.16, on which these theorems rest, a difficulty may arise with a limit point $\bar{x}$ for which $\mathcal{E}^-(\bar{x}) \neq \varnothing$, which is likely to be on a kink of H. When an iterate x_k is close to such an $\bar{x}$ and $i \in \mathcal{F}(x_k)$ say (by symmetry, the reasoning is the same if $i \in \mathcal{G}(x_k)$), the system (2.10) gives an information on the variation of F_i at x_k along d_k (through the equation $F_i(x_k) + F_i'(x_k)d_k = 0$) but nothing is said on the variation of G_i along the same direction (since $G_i(x_k) + G_i'(x_k)d_k$ may take any value), while an information on $G_i'(x_k)d_k$ may also be necessary when the linesearch at x_k explores the two sides of the kink. It happens, actually, that relaxing the equality $F_i(x_k) + F_i'(x_k)d_k = 0$ into the inequality $F_i(x_k) + F_i'(x_k)d_k \geqslant 0$ and adding the inequality $G_i(x_k) + G_i'(x_k)d_k \geqslant 0$ suffice to complete the proof (see its points 4.1.2 and 4.2.2), while keeping the descent property (see corollary 2.9).

We first present in section 2.3.1 the *exact* version (2.12) of the direction described in the previous paragraph and discuss its links with other directions. Next, we analyze its descent property in section 2.3.2 and exhibit the *inexact* version (2.22) of the direction, which also enjoys the descent property.

2.3.1 Directions

Based on the previous discussion, we introduce a device that is able to measure the proximity to a point $\bar{x}$ on a kink of H that is due to an index in $\mathcal{E}^-(\bar{x})$ (rather mysteriously, the proximity to a point $\bar{x}$ on a kink due to an index in $\mathcal{E}^{0+}(\bar{x})$ is not troublesome). Let $\tau \in (0,\infty)$ be the *kink tolerance*, used to detect such a proximity (normally τ should be small, but we want to be rather general at this stage of the presentation) and define the index set

$$\mathcal{E}_\tau^-(x) := \{i : F_i(x) < 0,\ G_i(x) < 0,\ |F_i(x) - G_i(x)| < \tau\}. \tag{2.11a}$$

We also define

$$\mathcal{E}_0^-(x) := \cap_{\tau>0}\mathcal{E}_\tau^-(x) = \{i : F_i(x) = G_i(x) < 0\} = \mathcal{E}^-(x), \tag{2.11b}$$

$$\mathcal{E}_\infty^-(x) := \cup_{\tau>0}\mathcal{E}_\tau^-(x) = \{i : F_i(x) < 0,\ G_i(x) < 0\}. \tag{2.11c}$$

Note that the set $\mathcal{E}_\tau^-(x)$ is expandinging with τ, meaning that $\mathcal{E}_{\tau_1}^-(x) \subseteq \mathcal{E}_{\tau_2}^-(x)$ when $0 \leqslant \tau_1 \leqslant \tau_2 \leqslant \infty$.

A direction d is said to be a *secured PNM direction* if it satisfies

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in E_F(x) \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in E_G(x) \\ F_i(x) + F_i'(x)d \geqslant 0 & \text{if } i \in I(x) \\ G_i(x) + G_i'(x)d \geqslant 0 & \text{if } i \in I(x), \end{cases} \tag{2.12}$$

where we have used the following index sets:

$$E_F(x) := \big[\mathcal{F}(x) \setminus \mathcal{E}_\tau^-(x)\big] \cup \mathcal{E}_\mathcal{F}^{0+}(x), \tag{2.13a}$$

$$E_G(x) := \big[\mathcal{G}(x) \setminus \mathcal{E}_\tau^-(x)\big] \cup \mathcal{E}_\mathcal{G}^{0+}(x), \tag{2.13b}$$

$$I(x) := \mathcal{E}_\tau^-(x), \tag{2.13c}$$

in which $\tau \in [0, \infty]$ and $(\mathcal{E}_\mathcal{F}^{0+}(x), \mathcal{E}_\mathcal{G}^{0+}(x))$ is some partition of $\mathcal{E}^{0+}(x)$ (both can depend on x). These index sets will be continually used in the sequel and it is important to observe that they form a partition of $[1:n]$, which is claimed by the next lemma. As a consequence of this lemma, the system (2.12) has $|E_F(x)| + |E_G(x)| = n - |I(x)|$ equalities and $2|I(x)|$ inequalities.

Lemma 2.6 ($(\boldsymbol{E_F}, \boldsymbol{E_G}, \boldsymbol{I})$ partition) *The triplet*

$$\big(E_F(x),\ E_G(x),\ I(x)\big) \tag{2.14}$$

forms a partition of $[1:n]$.

Proof. Observe first that the triplet covers $[1:n]$:

$$\begin{aligned} & E_F(x) \cup E_G(x) \cup I(x) \\ & \quad = \big(\mathcal{F}(x) \cup \mathcal{E}_\mathcal{F}^{0+}(x)\big) \cup \big(\mathcal{G}(x) \cup \mathcal{E}_\mathcal{G}^{0+}(x)\big) \cup \mathcal{E}_\tau^-(x) \qquad [(2.13)] \\ & \quad \supseteq \mathcal{F}(x) \cup \mathcal{G}(x) \cup \mathcal{E}^{0+}(x) \cup \mathcal{E}^-(x) \qquad [\mathcal{E}_\mathcal{F}^{0+}(x) \cup \mathcal{E}_\mathcal{G}^{0+}(x) = \mathcal{E}^{0+}(x),\ \mathcal{E}_\tau^-(x) \supseteq \mathcal{E}^-(x)] \\ & \quad = [1:n] \qquad [\mathcal{E}^{0+}(x) \cup \mathcal{E}^-(x) = \mathcal{E}(x) \text{ and } \mathcal{E}(x) \cup \mathcal{F}(x) \cup \mathcal{G}(x) = [1:n]]. \end{aligned}$$

To conclude, it suffices to observe that the sets of the triplet are two by two disjoint:

- if $i \in E_F(x)$, then, $i \notin E_G(x)$, because $\mathcal{F}(x) \cap (\mathcal{G}(x) \cup \mathcal{E}^{0+}(x)) = \varnothing$ and $\mathcal{E}_\mathcal{F}^{0+}(x) \cap (\mathcal{G}(x) \cup \mathcal{E}_\mathcal{G}^{0+}(x)) = \varnothing$;
- if $i \in E_F(x)$ then $i \notin I(x)$, since $(\mathcal{F}(x) \setminus \mathcal{E}_\tau^-(x)) \cap \mathcal{E}_\tau^-(x) = \varnothing$ and $\mathcal{E}_\mathcal{G}^{0+}(x) \cap \mathcal{E}_\tau^-(x) = \varnothing$;
- if $i \in E_G(x)$ then $i \notin I(x)$ for a similar reason as in the previous case (switch F and G). □

In some discussions, we have found convenient to detect the index set to which a particular index $i \in [1:n]$ belongs by looking at the position of $(F_i(x), G_i(x))$ in the graph of figure 2.1.

By taking a value of τ close to zero, the number of inequalities in (2.12) should be small and the computation of the direction should be inexpensive. Our proof of global convergence (theorem 3.16 and theorem 3.18) requires to have $\tau > 0$, however. Then, the

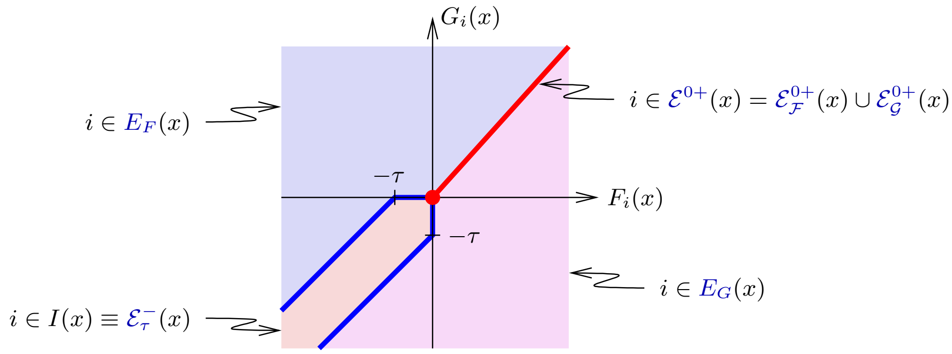

Figure 2.1: The pair $(F_i(x), G_i(x))$ determines in which of the index sets $E_F(x)$, $E_G(x)$, or $I(x)$, i belongs. The nondifferentiability of H_i can only occur on the main diagonal, at points x for which $F_i(x) = G_i(x)$. Nevertheless the *secured PNM direction* (2.12) carefully deals with points that are *near* an x such that $F_i(x) = G_i(x) < 0$, those in the tube in the left-bottom part of the picture (it then replaces one equality defining the plain NM direction (2.1) by a pair of inequalities). The width of this tube is controlled by the *kink tolerence* $\tau > 0$.

set $\mathcal{E}^-_\tau(x)$ is stable with respect to (or unchanged by) a small perturbation of x, which is not the case of $\mathcal{E}^-_0(x)$, so that the direction d defined by (2.12) is better adapted to the floating point arithmetic context than the direction (2.15) below, which corresponds to $\tau = 0$. Let us examine this last case.

By setting $\tau = 0$, at the left extreme point of the interval $[0, \infty]$, one has $\mathcal{E}^-_0(x) = \mathcal{E}^-(x)$, $\mathcal{F}(x) \setminus \mathcal{E}^-_0(x) = \mathcal{F}(x)$, and $\mathcal{G}(x) \setminus \mathcal{E}^-_0(x) = \mathcal{G}(x)$, so that the system (2.12) becomes

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x) \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) \\ F_i(x) + F_i'(x)d \geqslant 0 & \text{if } i \in \mathcal{E}^-(x) \\ G_i(x) + G_i'(x)d \geqslant 0 & \text{if } i \in \mathcal{E}^-(x). \end{cases} \tag{2.15}$$

This is the system (2.10) defining the plain PNM direction.

By setting $\tau = \infty$, at the right extreme point of the interval $[0, \infty]$, one has $\mathcal{F}(x) \setminus \mathcal{E}^-_\infty(x) = \{i : F_i(x) < G_i(x),\ G_i(x) \geqslant 0\}$, $\mathcal{G}(x) \setminus \mathcal{E}^-_\infty(x) = \{i : G_i(x) < F_i(x),\ F_i(x) \geqslant 0\}$, so that the system (2.12) becomes

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \{i : F_i(x) < G_i(x),\ G_i(x) \geqslant 0\} \cup \mathcal{E}^{0+}_{\mathcal{F}}(x) \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \{i : G_i(x) < F_i(x),\ F_i(x) \geqslant 0\} \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) \\ F_i(x) + F_i'(x)d \geqslant 0 & \text{if } i \in \mathcal{E}^-_\infty(x) \\ G_i(x) + G_i'(x)d \geqslant 0 & \text{if } i \in \mathcal{E}^-_\infty(x). \end{cases} \tag{2.16}$$

This system can be viewed as a relaxation of the following mixed LCP

$$\begin{cases} F_i(x) + F_i'(x)d = 0, & \text{if } i \in \{i : F_i(x) < G_i(x),\ G_i(x) \geqslant 0\} \cup \mathcal{E}^{0+}_{\mathcal{F}}(x) \\ G_i(x) + G_i'(x)d = 0, & \text{if } i \in \{i : G_i(x) < F_i(x),\ F_i(x) \geqslant 0\} \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) \\ 0 \leqslant (F(x) + F'(x)d)_{\mathcal{E}^-_\infty(x)} \perp (G(x) + G'(x)d)_{\mathcal{E}^-_\infty(x)} \geqslant 0, \end{cases}$$

which has an orthogonality condition that is not present in (2.16). This last system has some similarities with the system (1.12), obtained in [69] using other considerations, but is also rather different.

2.3.2 Descent property

The computation of a secured PNM direction satisfying (2.12), can be much more time consuming than solving the linear system (2.1) as required by the plain Newton-min direction. This is due to the presence of inequalities in the system (2.12). It is therefore tempting to see whether it is possible to design a criterion allowing an algorithm to take as often as possible the plain NM direction. This is the idea supporting the hybrid algorithm defined in section 2.4 and the first steps towards that algorithm are done in the present section: we focus on the design of such criterion and on its validation.

Around a solution, the plain NM direction is known to be appropriate because it yields fast convergence [56, 58], while this might not be the case far from a solution because it may fail to be a descent direction of the least-square merit function θ defined in (1.4); see counter-example 2.4. This observation speaks for a criterion based on the directional derivative of θ. Taking some safeguard, it could be appropriate to accept the plain NM direction d when it satisfies the inequality

$$\theta'(x;d) \leqslant -2(1-\eta)\,\theta(x) \tag{2.17}$$

where η is some constant in $[0,1)$. This inequality is natural since it is satisfied with $\eta = 0$ when d is the Newton direction on a smooth function H and θ is the map $x \mapsto \frac{1}{2}\|H(x)\|_2^2$. We have not been able to prove a global convergence result in the style of theorem 3.18 below with such a simple criterion.

For an arbitrary direction $d \in \mathbb{R}^n$, proposition 2.7 below will show, however, that

$$\sum_{i\in[1:n]} H_i(x)H_i'(x,d) = \theta'(x;d) \leqslant -\sum_{i\in[1:n]} (1-\rho_i(x,d))\,H_i(x)^2, \tag{2.18}$$

provided the $\rho_i(x,d)$'s are the unsigned values defined by formula (2.20) below. We shall see in corollary 2.9 that $\rho_i(x,d) \leqslant 0$ for the secured PNM direction (2.12), so that the inequality (2.17) with $\eta = 0$ follows from (2.18) for that direction. As a result, the secured PNM direction is a descent direction of θ at x (corollary 2.9).

The criterion for accepting an arbitrary direction d in the linesearch will be that the right-hand side of (2.18) is less than the right-hand side of (2.17), namely

$$-\sum_{i\in[1:n]} (1-\rho_i(x,d))\,H_i(x)^2 \leqslant -2(1-\eta)\,\theta(x), \tag{2.19a}$$

where η is some constant in $[0,1)$. From the expressions (1.3b) of H and (1.4) of θ, we see that this criterion simplifies into

$$\frac{1}{2}\sum_{i\in[1:n]} \rho_i(x,d)\,H_i(x)^2 \leqslant \eta\,\theta(x). \tag{2.19b}$$

The acceptation criterion (2.19) is more demanding than (2.17) since, thanks to (2.18), it implies (2.17). We see that the contributions of the terms in the sum in the left-hand side of (2.19) can be compensated by each other: the negativity of the directional derivative $\theta'(x;d)$ can be obtained by some negative terms in this sum, despite the positivity of other terms. This flexibility will allow the hybrid algorithm of section 2.4 to accept very often the plain NM direction. The important point is that this criterion happens to be sufficient to get the global convergence of theorem 3.18, because it is the left-hand side of this inequality that appears in its proof (see the one of theorem 3.16).

In the rest of this section, we focus on the proof of the inequality (2.18) and on its ability to detect descent directions. First, let us define the quantities $\rho_i(x,d)$ appearing in the right-hand side of (2.18). Let $x \in \mathbb{R}^n$ be an arbitrary point and $d \in \mathbb{R}^n$ be an arbitrary direction. We define $\rho_i(x,d)$ by

$$\rho_i(x,d) := \begin{cases} \frac{F_i(x)+F_i'(x)d}{F_i(x)} & \text{if } i \in E_F(x) \text{ and } F_i(x) \neq 0 \\ 0 & \text{if } i \in E_F(x) \text{ and } F_i(x) = 0 \\ \frac{G_i(x)+G_i'(x)d}{G_i(x)} & \text{if } i \in E_G(x) \text{ and } G_i(x) \neq 0 \\ 0 & \text{if } i \in E_G(x) \text{ and } G_i(x) = 0 \\ \max\left(\frac{F_i(x)+F_i'(x)d}{F_i(x)}, \frac{G_i(x)+G_i'(x)d}{G_i(x)}\right) & \text{if } i \in I(x), \end{cases} \tag{2.20}$$

where the partition $(E_F(x), E_G(x), I(x))$ of $[1\!:\!n]$ has been defined in (2.13) (hence, the five groups of indices in (2.20) also form a partition of $[1\!:\!n]$). The zero value given to $\rho_i(x,d)$ when $F_i(x)=0$ or $G_i(x)=0$ allows us to simplify the statement of corollary 2.9 below but, as we shall see, an arbitrary value could have been given instead, since this one does not occur in the calculations that follow. For $i \notin I(x)$, $\rho_i(x,d)$ measures the signed proximity of $F_i(x)+F_i'(x)d$ or $G_i(x)+G_i'(x)d$ to zero, compared to $F_i(x)$ or $G_i(x)$, respectively. For $i \in I(x)$, a positive $\rho_i(x,d)$ measures the relative failure in the realization of the positivity of both $F_i(x)+F_i'(x)d$ and $G_i(x)+G_i'(x)d$; if $\rho_i(x,d)$ is nonpositive, both $F_i(x)+F_i'(x)d$ and $G_i(x)+G_i'(x)d$ are nonnegative.

Let us stress the fact that the $\rho_i(x,d)$'s given by (2.20) are not necessarily less than one and such a restriction on d is not imposed in the next proposition. Hence, the formula (2.18) does not give an upper bound of $\theta'(x;d)$ as a sum of nonpositive terms and does not imply the negativity of that directional derivative.

Proposition 2.7 (overestimation of $\boldsymbol{\theta'(x;d)}$) *Let $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$, H be the function defined by* (1.3b)*, and the $\rho_i(x,d)$'s be defined by* (2.20)*. Then* (2.18) *holds.*

Proof. Thanks to (2.5), (1.3b), and (1.7), the directional derivative of θ at x in the direction d has the following value

$$\begin{aligned} \theta'(x;d) = {} & F_{\mathcal{F}(x)}(x)^\mathsf{T} F'_{\mathcal{F}(x)}(x)d + G_{\mathcal{G}(x)}(x)^\mathsf{T} G'_{\mathcal{G}(x)}(x)d \\ & + F_{\mathcal{E}(x)}(x)^\mathsf{T} \min\big(F'_{\mathcal{E}(x)}(x)d, G'_{\mathcal{E}(x)}(x)d\big). \end{aligned}$$

Let us show that

$$\forall\, i \in \mathcal{F}(x): \quad F_i(x)F_i'(x)d \leqslant -(1-\rho_i)F_i(x)^2, \tag{2.21a}$$

$$\forall\, i \in \mathcal{G}(x): \quad G_i(x)G_i'(x)d \leqslant -(1-\rho_i)G_i(x)^2, \tag{2.21b}$$

where we have adopted the simplification $\rho_i = \rho_i(x,d)$. Indeed, let $i \in \mathcal{F}(x) \cup \mathcal{G}(x)$.

- If $i \in \mathcal{F}(x) \setminus \mathcal{E}_\tau^-(x)$ and $F_i(x) \neq 0$, $(2.20)_1$ gives $F_i'(x)d = -(1-\rho_i)F_i(x)$, hence (2.21a) by multiplying both sides of this identity by $F_i(x)$.
- If $i \in \mathcal{F}(x)$ and $F_i(x) = 0$, (2.21a) is clearly satisfied with equality.
- If $i \in \mathcal{F}(x) \cap \mathcal{E}_\tau^-(x)$, $(2.20)_5$ gives $-F_i'(x)d \leqslant (1-\rho_i)F_i(x)$, hence (2.21a) by multiplying both sides of this inequality by $-F_i(x) > 0$.
- If $i \in \mathcal{G}(x)$, (2.21b) follows like in the three previous cases, by using G_i instead of F_i and $(2.20)_3$ instead of $(2.20)_1$.

Now using (2.21) and $F_i(x) = G_i(x)$ for $i \in \mathcal{E}(x)$, we get

$$\begin{aligned}\theta'(x;d) \leqslant & -\sum_{i\in\mathcal{F}(x)} (1-\rho_i)F_i(x)^2 - \sum_{i\in\mathcal{G}(x)} (1-\rho_i)G_i(x)^2 - \sum_{i\in\mathcal{E}(x)} (1-\rho_i)F_i(x)^2 \\ & + \sum_{i\in\mathcal{E}(x)} F_i(x)\min\big((1-\rho_i)F_i(x) + F_i'(x)d, (1-\rho_i)G_i(x) + G_i'(x)d\big).\end{aligned}$$

Therefore, to get (2.18), it suffices to show that the last term in the right-hand side of the previous inequality is nonpositive (we show that it vanishes).

- If $i \in \mathcal{E}^0(x)$, then $F_i(x) = G_i(x) = 0$ and the corresponding term vanishes.
- If $i \in \mathcal{E}^+(x) = (\mathcal{E}_\mathcal{F}^{0+}(x) \cup \mathcal{E}_\mathcal{G}^{0+}(x)) \setminus \mathcal{E}^0(x)$, then the arguments of the minimum function vanish by the definition of ρ_i in $(2.20)_1$ and $(2.20)_3$.
- If $i \in \mathcal{E}^-(x) = \mathcal{E}_0^-(x) \subseteq \mathcal{E}_\tau^-(x)$, then the minimum vanishes by $(2.20)_5$. $\square$

Counterexample 2.8 *The inequality* (2.17) *may be strict, even for the plain NM direction* (2.1) *for the LCP* (1.2). Consider indeed the LCP (1.2) with $n = 1$, $M = 1/2$, and $q = -1/4$. Suppose that $x = -1$ and that $\tau = 1/2$. Since $x < y := Mx + q = -3/4$, the single index 1 is in $\mathcal{G}(x)$ and the plain NM direction (2.1) solves $x + d = 0$, hence $d = 1$. Since both x and y are negative and $|x - y| = 1/4 < \tau$, $1 \in \mathcal{E}_\tau^-(x)$. Therefore,

$$\rho_1(x,d) = \max((x+d)/x, (y+Md)/y) = \max(0, 1/3) = 1/3.$$

We conclude that $\theta'(x;d) = xd = -1$ is strictly less that $-(1-\rho_1(x,d))H(x)^2 = -2/3$. $\square$

Corollary 2.9 (descent direction with (2.12)) *Suppose that d satisfies* (2.12) *for some $\tau \in [0,\infty]$. Then, the $\rho_i(x,d)$'s defined by* (2.20) *are nonpositive and, consequently,* (2.17) *and* (2.19) *hold with $\eta = 0$. In particular, d is a descent direction of θ at x.*

Proof. Suppose that d is defined by (2.12) at $x \in \mathbb{R}^n$. For $i \in E_F(x)$, $(2.12)_1$ shows that $F_i(x) + F_i'(x)d = 0$, so that $\rho_i(x,d) = 0$ by $(2.20)_1$ and $(2.20)_2$. Similarly, $\rho_i(x,d) = 0$ for $i \in E_G(x)$. For $i \in I(x)$, $(2.12)_3$ and $(2.12)_4$ show that $\rho_i(x,d) \leqslant 0$ by $(2.20)_5$. We have shown that the $\rho_i(x,d)$'s defined by (2.20) are nonpositive. Now, the inequality $\theta'(x;d) \leqslant -2\,\theta(x)$ follows immediately from (2.18), since the terms with $\rho_i(x,d)$ in factor can be discarded. For the same reason, (2.19) holds with $\eta = 0$. $\square$

As another illustration of the usefulness of proposition 2.7, consider an *inexact secured PNM direction* d, which, by definition, verifies, for some $\eta \geqslant 0$, the following inequalities:

$$
\begin{aligned}
F_i(x) + F_i'(x)d \leqslant \eta F_i(x), &\quad \forall\, i \in \mathcal{F}^{0+}(x) \cup \mathcal{E}_{\mathcal{F}}^{0+}(x), &\quad (2.22a)\\
\eta F_i(x) \leqslant F_i(x) + F_i'(x)d, &\quad \forall\, i \in \mathcal{F}^{-}(x) \cup \mathcal{E}_{\tau}^{-}(x), &\quad (2.22b)\\
G_i(x) + G_i'(x)d \leqslant \eta G_i(x), &\quad \forall\, i \in \mathcal{G}^{0+}(x) \cup \mathcal{E}_{\mathcal{G}}^{0+}(x), &\quad (2.22c)\\
\eta G_i(x) \leqslant G_i(x) + G_i'(x)d, &\quad \forall\, i \in \mathcal{G}^{-}(x) \cup \mathcal{E}_{\tau}^{-}(x), &\quad (2.22d)
\end{aligned}
$$

where the index sets $\mathcal{F}^{-}(x)$, $\mathcal{F}^{0+}(x)$, $\mathcal{G}^{-}(x)$, and $\mathcal{G}^{0+}(x)$ are defined by

$$
\begin{aligned}
\mathcal{F}^{-}(x) &:= \{i \in \mathcal{F}(x) : F_i(x) < 0\}, & \mathcal{F}^{0+}(x) &:= \{i \in \mathcal{F}(x) : F_i(x) \geqslant 0\},\\
\mathcal{G}^{-}(x) &:= \{i \in \mathcal{G}(x) : G_i(x) < 0\}, & \mathcal{G}^{0+}(x) &:= \{i \in \mathcal{G}(x) : G_i(x) \geqslant 0\}.
\end{aligned}
$$

It is simple verification to see that these conditions (2.22) are satisfied by a secured PNM direction, i.e., a direction d verifying (2.12). The conditions (2.22) are not very demanding, in particular because there is no equality to satisfy. During our exploration of the design of a criterion for accepting as often as possible a plain NM direction (2.1), the fact that it satisfied these conditions (2.22) was retained for a while, because they ensure the global convergence of section 3.2. Actually, as shown by the next corollary, an inexact secured PNM direction also satisfies the criterion (2.19), which is an indirect way of showing that it guarantees the global convergence results of section 3.2. Since (2.22) implies (2.19), the latter criterion is less demanding, more often verified, than the former, which is the reason why we have adopted the criterion (2.19) in section 2.4.

Corollary 2.10 (descent direction with (2.22)) *Suppose that d satisfies (2.22) for some $\tau \in [0, \infty]$ and $\eta \geqslant 0$. Then, the $\rho_i(x, d)$'s defined by (2.20) do not exceed η and, consequently, (2.17) and (2.19) hold with the given η. In particular, if $\eta \in [0, 1)$, d is a descent direction of θ at x.*

Proof. Suppose that d is satisfies (2.22) at $x \in \mathbb{R}^n$ for some $\tau \in [0, \infty]$ and $\eta \geqslant 0$.

There holds

$$
E_F(x) = \mathcal{F}^{0+}(x) \cup \mathcal{E}_{\mathcal{F}}^{0+}(x) \cup [\mathcal{F}^{-}(x) \setminus \mathcal{E}_{\tau}^{-}(x)].
$$

If $i \in \mathcal{F}^{0+}(x) \cup \mathcal{E}_{\mathcal{F}}^{0+}(x)$ and $F_i(x) \neq 0$, $(2.20)_1$, (2.22a), and the positivity of $F_i(x)$ give $\rho_i(x, d) \leqslant \eta$. If $i \in \mathcal{F}^{0+}(x) \cup \mathcal{E}_{\mathcal{F}}^{0+}(x)$ and $F_i(x) = 0$, $(2.20)_2$ and $\eta \geqslant 0$ show that $\rho_i(x, d) \leqslant \eta$. If $i \in \mathcal{F}^{-}(x) \setminus \mathcal{E}_{\tau}^{-}(x)$, $(2.20)_1$, (2.22b), and the negativity of $F_i(x)$ give again $\rho_i(x, d) \leqslant \eta$. We have shown that $\rho_i(x, d) \leqslant \eta$, for $i \in E_F(x)$.

For similar reasons, $\rho_i(x, d) \leqslant \eta$, for $i \in E_G(x)$.

Consider now the indices $i \in I(x) = \mathcal{E}_{\tau}^{-}(x)$. By (2.22b), (2.22d), and the negativity of $F_i(x)$ and $G_i(x)$, we get $(F_i(x) + F_i'(x)d)/F_i(x) \leqslant \eta$ and $(G_i(x) + G_i'(x)d)/G_i(x) \leqslant \eta$, so that $\rho_i(x, d) \leqslant \eta$ by $(2.20)_5$.

In conclusion, we have shown that

$$
\rho_i(x, d) \leqslant \eta, \quad \text{for } i \in [1\,{:}\,n].
$$

Therefore, the criterion (2.19) is satisfied. Finally, using (2.18), which is guaranteed by proposition 2.7, and (2.19), we get (2.17). □

2.4 The hybrid Newton-min algorithm

The directions presented in section 2.3 give rise to several algorithms that follow the same principles, which are gathered in the following generic algorithm. It is the global convergence of this generic algorithm that will be analyzed in section 3.2, and more particularly in theorem 3.16, in which an additional assumption is made on the computed directions. In this algorithm, the term "constant" means "independent of the iteration".

Algorithm 2.11 (generic NM algorithm) Let $\eta \in [0, 1)$ be the constant appearing in the acceptation criterion (2.19), let $\tau \in (0, \infty]$ be a constant used in the computation of the direction d, and let $\omega \in (0, 1)$ and $\beta \in (0, 1)$ be the two constants used in the linesearch. Let x be the current iterate. The next iterate $x_+ \in \mathbb{R}^n$ is computed as follows.

1. *Stopping criterion.* If $\theta(x) = 0$, stop (then, x is a solution to (1.1)).
2. *Direction.* Compute a direction $d \in \mathbb{R}^n$ satisfying (2.19).
3. *Stepsize.* Set $\alpha := \beta^i$, where i is the smallest nonnegative integer such that

$$\theta(x + \alpha d) \leqslant (1-2\omega\alpha(1-\eta))\,\theta(x). \tag{2.23}$$

4. *New iterate.* $x_+ := x + \alpha d$.

The well-posedness of this algorithm is discussed below, after having presented two of its instances.

A first instance of the generic NM algorithm is the one that computes the direction d as the minimum norm solution to (2.12).

Algorithm 2.12 (PNM algorithm) It is the instance of algorithm 2.11, in which the direction d in step 2 is computed as follows: choose some partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$ and then compute the solution to the following problem

$$\min\,\{\|d\| : d \text{ satisfies } (2.12)\}. \tag{2.24}$$

The partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$ intervenes in the definition of the index sets $E_F(x)$, $E_G(x)$, and $I(x)$ by (2.13). The norm $\|\cdot\|$ in (2.24) need not be specified for the convergence of the algorithm, but in practice, we use the Euclidean norm, so that the problem is quadratic. Since the solution d to (2.24) satisfies (2.12), it satisfies (2.19) (corollary 2.9), which shows that algorithm 2.12 is indeed an instance of algorithm 2.11.

As already discussed at the beginning of section 2.3.2, the *hybrid Newton-min* (HNM) *algorithm* presented now aims at reducing the cost of the computation of a descent direction of algorithm 2.12 by accepting the plain NM direction (2.1) as soon as it satisfies the criterion (2.19) ; if this criterion is not satisfied, a secured PNM direction, hence satisfying (2.12), is computed. As we shall see, in addition to minimizing the cost of the iteration, this approach also ensures global convergence (section 3.2).

Algorithm 2.13 (HNM algorithm) It is the instance of algorithm 2.11, in which the direction d in step 2 is computed as follows: the algorithm first computes a plain NM direction $d \in \mathbb{R}^n$, hence satisfying (2.1); if (2.19) does not hold with that d, it chooses some partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$ and recomputes the direction d as the solution to (2.24).

These algorithms are rather standard in their form. Only the computation of the direction in step 2, whose conception has been progressively introduced above, makes exception. Let us give some more comments.

1. There are implicit assumptions in step 2, which will have to be clarified in the results on these algorithms, namely
   - it assumes that (2.1) has always a solution, which may not be the case if the Jacobian of this linear system is singular;
   - similarly, it assumes that problem (2.24) has a solution when desired, which may not be the case if the affine system (2.12) is infeasible; a rather weak condition guaranteeing the feasibility of (2.12), for x near a limit point $\bar{x}$, is discussed in section 3.1.
2. If not empty, the polyhedron defined by (2.12) may be unbounded, which raises some difficulty in the convergence proof of section 3.2. For this reason, in (2.24), we take the option of taking a minimum norm direction in that polyhedron.
3. The directions computed in step 2, if any, are necessarily descent directions of θ at x. This is because they satisfy (2.19) with $\eta < 1$, hence (2.17) with the same $\eta < 1$, implying that $\theta'(x; d) < 0$ when x is not a solution. As a result, the linesearch in step 3 is able to compute a stepsize $\alpha > 0$ in a finite number of trials, when x is not a solution to the complementarity problem (this is guaranteed by step 1).
4. Condition (2.23) derives from the standard Armijo inequality [3, 31, 14]

$$\theta(x + \alpha d) \leqslant \theta(x) + \omega\alpha\theta'(x; d),$$

   in which the negative upper bound $-2(1-\eta)\,\theta(x)$ of $\theta'(x; d)$ given by (2.17) has substituted the directional derivative.

3 Algorithm analysis

This section starts with establishing verifiable conditions at a given point $\bar{x} \in \mathbb{R}^n$ to ensure that the system (2.12), defining the secured PNM direction d, has a solution when x is near $\bar{x}$ (section 3.1). It is then shown that these conditions also ensure that these directions can be chosen such that they remain bounded for x near $\bar{x}$ (section 3.1.3). This boundedness property is useful for establishing the global convergence result (section 3.2).

3.1 Regularity

3.1.1 Regularity of a point

Let $\bar{x} \in \mathbb{R}^n$ be a point that is not necessarily a solution to the complementarity problem (1.1). Our vehicle for highlighting conditions ensuring the solubility of the affine system (2.12), when x is near $\bar{x}$, is the Mangasarian-Fromovitz constraint qualification (MFCQ) [63], which will be reinforced to deal with the present special situation.

It is well known that the system (2.12) has a solution d when MFCQ holds at x. This "constraint qualification" has several expressions, one of them being that

$$\begin{array}{l}\sum_{i\in E_F(x)} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(x)} \beta_i \nabla G_i(x) + \sum_{i\in I(x)} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0 \\ \text{and } (\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0,\end{array} \tag{3.1}$$

Another equivalent version, which will be used occasionally, reads

$$\begin{array}{l}\text{The map } d \in \mathbb{R}^n \mapsto (F'_{E_F(x)}(x)d, G'_{E_G(x)}(x)d) \in \mathbb{R}^{|E_F(x)|} \times \mathbb{R}^{|E_G(x)|} \text{ is surjective} \\ \text{and there is a direction } d \in \mathbb{R}^n \text{ such that } F'_{E_F(x)}(x)d = 0,\ G'_{E_G(x)}(x)d = 0, \\ F'_{I(x)}(x)d > 0, \text{ and } G'_{I(x)}(x)d > 0.\end{array} \tag{3.2}$$

Through the index sets $E_F(x)$ and $E_G(x)$, condition (3.1) involves a partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, which can be chosen arbitrarily by the algorithm at an $x \neq \bar{x}$. This fact will contribute to the complexity of the analysis. We quote this regularity condition at x in the next proposition.

Proposition 3.1 (punctual regularity) *Suppose that F and G are differentiable at some point $x \in \mathbb{R}^n$ and that (3.1) holds at that point. Then, the system (2.12) has a solution $d \in \mathbb{R}^n$.*

What is actually desired for future use is an MFCQ-like condition at $\bar{x}$ implying that (3.1) holds for x near $\bar{x}$, and therefore that (2.12) has a solution for x near $\bar{x}$. In other words, we want to have the implication:

$$\text{Strengthening of (3.1) at } x = \bar{x} \qquad \Longrightarrow \qquad \text{(3.1) at } x \text{ near } \bar{x}, \tag{3.3}$$

in which the "strengthening of (3.1) at $x = \bar{x}$" has to be specified. Instead of giving now the strengthened version of (3.1) that is used in the sequel, we prefer introducing it by showing what motivates its definition. We structure the discussion in five stages.

Discussion 3.2 1. The reason why (3.1) must be strengthened to have the implication in (3.3) comes from the change in the index sets with x. Suppose indeed that only (3.1) holds at $x = \bar{x}$ (i.e., no strengthening). It is well known that the implication in (3.1) is insensitive to small perturbations in the gradients $\nabla F_i(x)$ and $\nabla G_i(x)$ in its premise (this can be deduced from its equivalent form (3.2)). Therefore, if we assume the continuity of the derivatives F' and G' at $\bar{x}$ and if x is neat $\bar{x}$, it follows from (3.1) at $x = \bar{x}$ that

$$\begin{array}{l}\sum_{i\in E_F(\bar{x})} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(\bar{x})} \beta_i \nabla G_i(x) + \sum_{i\in I(\bar{x})} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0 \\ \text{and } (\alpha_{I(\bar{x})}, \beta_{I(\bar{x})}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0.\end{array} \tag{3.4}$$

As announced, this implication may no longer be true if the index sets are evaluated at x instead of $\bar{x}$. In other words, it may occur that (3.4) does not imply (3.1), even if x is very close to $\bar{x}$. Let us take a closer look at this difficulty.

2. Somehow, the index set $I(\bar{x})$ defined in (2.13c) or (2.11a) raises no difficulty, in part because of the following observation.

Lemma 3.3 ($I(\bar{x})$ and $I(x)$) *Suppose that F and G are continuous at $\bar{x}$, that $\tau \in (0, \infty]$, and that x is near $\bar{x}$. Then*

$$I(\bar{x}) \subseteq I(x). \tag{3.5}$$

Proof. This follows immediately from the assumptions and the strict inequalities defining $I(x)$. $\square$

Using (3.5), we see that (3.4) implies

$$\begin{array}{l}\sum_{i\in E_F(\bar{x})\setminus I(x)} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(\bar{x})\setminus I(x)} \beta_i \nabla G_i(x) + \sum_{i\in I(x)} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0 \\ \text{and } (\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0,\end{array}$$

which resembles more (3.1), at least by one index set. This implication is valid, because some unsigned multipliers α_i and β_i in the premise of (3.4) may have become signed multipliers in the premise of the last implication. This implication is therefore more often satisfied than (3.4). Nevertheless, one cannot go further in the direction of (3.1) without strengthening (3.4).

3. The difficulty arises actually from the indices $i \in \mathcal{E}^{0+}(\bar{x}) := \{i \in [1\!:\!n] : F_i(\bar{x}) = G_i(\bar{x}) \geqslant 0\}$, which, for x arbitrarily close to $\bar{x}$, can be in any of the index sets defined at x that form a partition of $[1\!:\!n]$, namely $E_F(x)$, $E_G(x)$, and $I(x)$. If the considered index i is in $I(x)$, there is no difficulty, for the same reason as the one mentioned above: some unsigned multipliers in (3.4) become signed multipliers in (3.4) when $\bar{x}$ is substituted by the neighboring point x. Now, an index i belonging to $\mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}) \subseteq E_G(\bar{x})$ and $\mathcal{F}(x) \setminus I(x) \subseteq E_F(x)$, which is compatible, is more troublesome since the associated multiplier β_i in (3.4) would become a multiplier α_i in (3.4) with $(E_F(\bar{x}), E_G(\bar{x}), I(\bar{x}))$ changed into $(E_F(x), E_G(x), I(x))$. To clarify the difficulty, note that the indices $i \in E_F(x)$ involve $\nabla F_i(x)$, while the indices $i \in E_G(\bar{x})$ involve $\nabla G_i(x)$. Then, to get the implication (3.3), one would have to strengthen much (3.1) at $x = \bar{x}$ by adding the terms

$$\sum_{i\in E_F(\bar{x})} \tilde{\alpha}_i \nabla G_i(\bar{x}) + \sum_{i\in E_G(\bar{x})} \tilde{\beta}_i \nabla F_i(\bar{x})$$

in the identity in the premise of the implication, with unsigned $\tilde{\alpha}_i$'s and $\tilde{\beta}_i$'s. This is a too important strengthening, and it can be avoided as follows.

Instead of considering all neighboring points x to $\bar{x}$, we only consider those for which

$$E_F(x) \subseteq E_F(\bar{x}) \qquad \text{and} \qquad E_G(x) \subseteq E_G(\bar{x}). \tag{3.6}$$

By the partition property of lemma 2.6 and (3.6), one has

$$\big(E_F(\bar{x}) \setminus E_F(x)\big) \cup \big(E_G(\bar{x}) \setminus E_G(x)\big) = I(x) \setminus I(\bar{x}). \tag{3.7}$$

In plain words, this identity shows that the indices in $E_F(\bar{x})$ that do not stay in $E_F(x)$ and those in $E_G(\bar{x})$ that do not stay in $E_G(x)$, when $\bar{x}$ becomes a neighboring point x satisfying (3.6), go into $I(x)$, which is not troublesome as we have seen above. To be more precise, we claim that, for a neighboring point x satisfying (3.6), one gets the implication (3.1) from the one in (3.4). Indeed, assume that the premise of (3.1) holds:

$$\textstyle\sum_{i\in E_F(x)} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(x)} \beta_i \nabla G_i(x) + \sum_{i\in I(x)} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0$$
and $(\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0$.

Then, by (3.7),

$$\textstyle\sum_{i\in E_F(\bar{x})} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(\bar{x})} \beta_i \nabla G_i(x) + \sum_{i\in I(\bar{x})} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0$$
and $(\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0$.

By lemma 3.3, $I(\bar{x}) \subseteq I(x)$, so that we certainly have

$$\textstyle\sum_{i\in E_F(\bar{x})} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(\bar{x})} \beta_i \nabla G_i(x) + \sum_{i\in I(\bar{x})} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0$$
and $(\alpha_{I(\bar{x})}, \beta_{I(\bar{x})}) \geqslant 0$.

Finally, by (3.4), one gets $(\alpha, \beta) = 0$.

4. The question now arises to know whether all points x near $\bar{x}$ are covered by the previous approach, which restricts the considered x's to those satisfying (3.6). To answer this question, we start with observing that both $E_F(\bar{x})$ and $E_G(\bar{x})$ depend on the choice of the partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$. We can then make the question more precise by reformulating it as the one asking whether, for all x near $\bar{x}$, one can find such a partition yielding (3.6). The answer is positive.

Lemma 3.4 (adapted partition of $\mathcal{E}^{0+}(\bar{x})$) *Suppose that F and G are continuous at $\bar{x}$, that $\tau \in (0, \infty]$, and that x is near $\bar{x}$. Let $(\mathcal{E}_{\mathcal{F}}(x), \mathcal{E}_{\mathcal{G}}(x))$ be any partition of $\mathcal{E}(x)$ such that*

$$\mathcal{E}^{0+}_{\mathcal{F}}(x) = \mathcal{E}^{0+}(x) \cap \mathcal{E}_{\mathcal{F}}(x) \qquad \textit{and} \qquad \mathcal{E}^{0+}_{\mathcal{G}}(x) = \mathcal{E}^{0+}(x) \cap \mathcal{E}_{\mathcal{G}}(x). \tag{3.8a}$$

Such a partition always exists. Define the partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ by

$$\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}) := \mathcal{E}^{0+}(\bar{x}) \cap \big[\mathcal{F}(x) \cup \mathcal{E}_{\mathcal{F}}(x)\big], \tag{3.8b}$$
$$\mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}) := \mathcal{E}^{0+}(\bar{x}) \cap \big[\mathcal{G}(x) \cup \mathcal{E}_{\mathcal{G}}(x)\big]. \tag{3.8c}$$

Then, the inclusions in (3.6) hold.

Proof. Note first that it is always possible to find a partition $(\mathcal{E}_{\mathcal{F}}(x), \mathcal{E}_{\mathcal{G}}(x))$ of $\mathcal{E}(x)$, such that (3.8a) holds. This is because $\mathcal{E}(x)$ is the disjoint union of $\mathcal{E}^{-}(x)$ and $\mathcal{E}^{0+}(x)$.

Hence, taking any partition $(\mathcal{E}^-_{\mathcal{F}}(x), \mathcal{E}^-_{\mathcal{G}}(x))$ of $\mathcal{E}^-(x)$ and setting $\mathcal{E}_{\mathcal{F}}(x) := \mathcal{E}^-_{\mathcal{F}}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x)$ and $\mathcal{E}_{\mathcal{G}}(x) := \mathcal{E}^-_{\mathcal{G}}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)$ yield the desired partition.

Note also that $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ defined by (3.8b) and (3.8c) is indeed a partition of $\mathcal{E}^{0+}(\bar{x})$. This is because $(\mathcal{F}(x) \cup \mathcal{E}_{\mathcal{F}}(x), \mathcal{G}(x) \cup \mathcal{E}_{\mathcal{G}}(x))$ is a partition of $[1\,{:}\,n]$.

To show that (3.6) holds, suppose that $i \in E_F(x) := \big[\mathcal{F}(x) \setminus I(x)\big] \cup \mathcal{E}^{0+}_{\mathcal{F}}(x)$; see (2.13) for the formula. We pursue by considering two cases.

- If $i \in \mathcal{F}(x) \setminus I(x)$, then, $i \in \mathcal{F}(\bar{x}) \cup \mathcal{E}(\bar{x})$ [because $x \in \mathcal{F}(x)$ is near $\bar{x}$] and $i \in \mathcal{F}(x) \setminus I(\bar{x})$ [by (3.5)]. We consider two subcases, exploiting the fact that $i \in \mathcal{F}(\bar{x}) \cup \mathcal{E}(\bar{x})$.
  - If $i \in \mathcal{F}(\bar{x})$, then $i \in \mathcal{F}(\bar{x}) \setminus I(\bar{x}) \subseteq E_F(\bar{x})$.
  - If $i \in \mathcal{E}(\bar{x})$, then one must have $i \in \mathcal{E}^{0+}(\bar{x})$ (since $i \notin I(\bar{x})$). Now, by (3.8b) and $i \in \mathcal{F}(x)$, it follows that $i \in \mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}) \subseteq E_F(\bar{x})$.
- If $i \in \mathcal{E}^{0+}_{\mathcal{F}}(x)$, then, by $(3.8\text{a})_1$, one must have $i \in \mathcal{E}^{0+}(x)$, hence $i \in \mathcal{E}^{0+}(\bar{x})$ when x is close to $\bar{x}$, and one must have $i \in \mathcal{E}_{\mathcal{F}}(x)$. By (3.8b), it follows that $i \in \mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}) \subseteq E_F(\bar{x})$.

In each case, we have shown that $i \in E_F(\bar{x})$, hence $E_F(x) \subseteq E_F(\bar{x})$.

The proof that $E_G(x) \subseteq E_G(\bar{x})$ works similarly by switching F and G, and using (3.8c) instead of (3.8b), and $(3.8\text{a})_2$ instead of $(3.8\text{a})_1$. □

5. Before stating the strengthened version of (3.1) at $x = \bar{x}$, we still have to clarify one point. Up to now, we have obtained an MFCQ-like condition at x from (3.4), by looking at the variation of $E_F(\bar{x})$, $E_G(\bar{x})$, and $I(\bar{x})$ when $\bar{x}$ becomes a neighboring point x. To complete the journey, we would like to know the range of sets in which $E_F(x)$, $E_G(x)$, and $I(x)$ can lie, when x is near $\bar{x}$. Define

$$I^\circ(\bar{x}) := \{i \in [1\,{:}\,n] : F_i(\bar{x}) \leqslant 0,\ G_i(\bar{x}) \leqslant 0,\ |F_i(\bar{x}) - G_i(\bar{x})| \leqslant \tau\}, \tag{3.9a}$$
$$E^\circ_F(\bar{x}) := E_F(\bar{x}) \setminus I^\circ(\bar{x}), \tag{3.9b}$$
$$E^\circ_G(\bar{x}) := E_G(\bar{x}) \setminus I^\circ(\bar{x}). \tag{3.9c}$$

Lemma 3.5 (range of $I(\cdot)$) *Suppose that F and G are continuous at $\bar{x}$, that $\tau \in (0, \infty]$, and that x is near $\bar{x}$. Then,*

$$I(\bar{x}) \subseteq I(x) \subseteq I^\circ(\bar{x}). \tag{3.10a}$$

Proof. The first inclusion is (3.5). For proving the second inclusion, assume that $i \notin I^\circ(\bar{x})$. Then, one of the nonstrict inequalities defining $I^\circ(\bar{x})$ is in the strict reverse sense. This one is also verified in the strict reverse sence by a neighboring x. Hence $i \notin I(x)$. □

Lemma 3.6 (**(E_F°, E_F°, I°) partition**) *The triplet* $\big(E_F^\circ(\bar{x}),\ E_G^\circ(\bar{x}),\ I^\circ(\bar{x})\big)$ *forms a partition of* $[1:n]$.

Proof. The union of the member sets cover $[1:n]$:

$$\begin{aligned}
&E_F^\circ(\bar{x}) \cup E_G^\circ(\bar{x}) \cup I^\circ(\bar{x})) \\
&\quad = \big(E_F(\bar{x}) \setminus I^\circ(\bar{x})\big) \cup \big(E_G(\bar{x}) \setminus I^\circ(\bar{x})\big) \cup I^\circ(\bar{x}) \qquad [(3.9b) \text{ and } (3.9c)] \\
&\quad = E_F(\bar{x}) \cup E_G(\bar{x}) \cup I^\circ(\bar{x}) \\
&\quad \supseteq E_F(\bar{x}) \cup E_G(\bar{x}) \cup I(\bar{x}) \qquad [(3.10a)] \\
&\quad = [1:n] \qquad [\text{lemma } 2.6].
\end{aligned}$$

Furthermore, the member sets are two by two disjoint. This is clear for $E_F^\circ(\bar{x})$ and $E_G^\circ(\bar{x})$ from the definitions (3.9b) and (3.9c), and since $E_F(\bar{x}) \cap E_G(\bar{x}) = \varnothing$ by lemma 2.6. It is also clear from the definitions (3.9b) and (3.9c) that $E_F^\circ(\bar{x}) \cap I^\circ(\bar{x}) = \varnothing$ and that $E_G^\circ(\bar{x}) \cap I^\circ(\bar{x}) = \varnothing$. □

Lemma 3.7 (**range of $E_F(\cdot)$, $E_G(\cdot)$**) *Suppose that F and G are continuous at $\bar{x}$, that $\tau \in (0,\infty]$, and that x is near $\bar{x}$. Suppose also that the partition $(\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x}), \mathcal{E}_{\mathcal{G}}^{0+}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ is derived from the partition $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ of $\mathcal{E}^{0+}(x)$ by the construction (3.8). Then,*

$$E_F^\circ(\bar{x}) \subseteq E_F(x) \subseteq E_F(\bar{x}) \subseteq E_F(x) \cup I(x), \tag{3.10b}$$
$$E_G^\circ(\bar{x}) \subseteq E_G(x) \subseteq E_G(\bar{x}) \subseteq E_G(x) \cup I(x), \tag{3.10c}$$

Proof. [(3.10b)] We know from lemma 3.4 that the construction (3.8) guarantees the median inclusion $E_F(x) \subseteq E_F(\bar{x})$.

Consider now the right-hand side inclusion $E_F(\bar{x}) \subseteq E_F(x) \cup I(x)$. Let $i \in E_F(\bar{x})$. By lemma 2.6, we only have to show that $i \notin E_G(x)$. In view of the definition (2.13a) of $E_F(\bar{x}) := \big[\mathcal{F}(\bar{x}) \setminus I(\bar{x})\big] \cup \mathcal{E}_{\mathcal{F}}^{0+}(\bar{x})$, we consider two cases.

- Suppose first that $i \in \mathcal{F}(\bar{x}) \setminus I(\bar{x})$. Then, $F_i(\bar{x}) < G_i(\bar{x})$. By the continuity of F and G, the proximity of x to $\bar{x}$, and the strict inequality, it follows that $i \in \mathcal{F}(x)$. Then, $i \notin \big[\mathcal{G}(x) \setminus I(x)\big] \cup \mathcal{E}_{\mathcal{G}}^{0+}(x) =: E_G(x)$.
- Suppose now that $i \in \mathcal{E}_{\mathcal{F}}^{0+}(\bar{x})$. Then, by the construction (3.8b), $i \in \mathcal{F}(x) \cup \mathcal{E}_{\mathcal{F}}(x)$. As a result, $i \notin \mathcal{G}(x) \setminus I(x)$ [since $i \notin \mathcal{G}(x)$] and $i \notin \mathcal{E}_{\mathcal{G}}^{0+}(x) := \mathcal{E}^{0+}(x) \cap \mathcal{E}_{\mathcal{G}}(x)$ [since $i \in \mathcal{E}_{\mathcal{F}}(x)$]. Hence, $i \notin E_G(x)$.

Consider finally the left-hand side inclusion $E_F^\circ(\bar{x}) \subseteq E_F(x)$. Using the just proven inclusion $E_F(\bar{x}) \subseteq E_F(x) \cup I(x)$ and lemma 2.6, we get $E_F(\bar{x}) \setminus I(x) \subseteq E_F(x)$. Now, since $I(x) \subseteq I^\circ(\bar{x})$ by (3.10a), one certainly has $E_F^\circ(\bar{x}) := E_F(\bar{x}) \setminus I^\circ(\bar{x}) \subseteq E_F(x)$.

[(3.10c)] The proof is similar to the one of (3.10b), by exchanging the roles of F and G. $\square$

Note that $E_F^\circ(\bar{x}) := E_F(\bar{x}) \setminus I^\circ(\bar{x})$ may be larger than $\mathcal{F}(\bar{x}) \setminus I^\circ(\bar{x})$, since the former set may contain the indices in the possibly nonempty set $\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x})$ and these indices are not in the latter set. Therefore, the inclusion $E_F^\circ(\bar{x}) \subseteq E_F(x)$ in (3.10b) is stronger than the inclusion $\mathcal{F}(\bar{x}) \setminus I^\circ(\bar{x}) \subseteq E_F(x)$, which follows from a straighforward argument. This stronger inclusion is due to the special partition $(\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x}), \mathcal{E}_{\mathcal{G}}^{0+}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$, which is derived from the partition $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ of $\mathcal{E}^{0+}(x)$ by the construction (3.8). It would not hold for disconnected partitions of $\mathcal{E}^{0+}(\bar{x})$ and $\mathcal{E}^{0+}(x)$.

We can now specify the regularity condition at $\bar{x}$ that we adopt. Let τ be fixed in $(0, \infty]$. The partition $(\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x}), \mathcal{E}_{\mathcal{G}}^{0+}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ that is specified at the beginning of the statement of the condition determines the index sets $E_F(\bar{x})$ and $E_G(\bar{x})$ defined by (2.13) at $x = \bar{x}$, and the index sets $E_F^\circ(\bar{x})$ and $E_G^\circ(\bar{x})$ defined by (3.9b) and (3.9c). Note also that the point of view adopted in this regularity definition is somehow reverse to the one in lemma 3.4 ; there, the partition is adapted to a given x near $\bar{x}$, in order to have the inclusions (3.6) ; here, all the possible partitions $(\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x}), \mathcal{E}_{\mathcal{G}}^{0+}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ are considered to be sure to have the inclusions (3.6), whatever the neighboring points x are.

Definition 3.8 (PNM regularity) For a given $\tau \in (0, \infty]$, a point $\bar{x} \in \mathbb{R}^n$ is said to be *PNM regular* (we also say that the complementarity problem (1.1) is *PNM regular* at $\bar{x} \in \mathbb{R}^n$), if F and G are differentiable at $\bar{x}$ and if for any partition $(\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x}), \mathcal{E}_{\mathcal{G}}^{0+}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$, and any partition (E_F, E_G, I) of $[1\!:\!n]$ such that $E_F^\circ(\bar{x}) \subseteq E_F \subseteq E_F(\bar{x})$, $E_G^\circ(\bar{x}) \subseteq E_G \subseteq E_G(\bar{x})$, and $I(\bar{x}) \subseteq I \subseteq I^\circ(\bar{x})$, there holds

$$\begin{aligned}&\textstyle\sum_{i\in E_F} \alpha_i \nabla F_i(\bar{x}) + \sum_{i\in E_G} \beta_i \nabla G_i(\bar{x}) + \sum_{i\in I} \big[\alpha_i \nabla F_i(\bar{x}) + \beta_i \nabla G_i(\bar{x})\big] = 0 \\ &\text{and } (\alpha_I, \beta_I) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0.\end{aligned} \tag{3.11}$$

$\square$

The relevance of this regularity condition derives from the following proposition, which shows that the PNM regularity of a point diffuses to the neighboring points (see also proposition 2.2).

Proposition 3.9 (diffusing PNM regularity) *Suppose that F and G are differentiable near some $\bar{x} \in \mathbb{R}^n$, that F' and G' are continuous at $\bar{x}$, and that $\bar{x}$ is PNM regular in the sense of definition 3.8. Then, for any x near $\bar{x}$ and any partition $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ of $\mathcal{E}^{0+}(x)$, the MFCQ (3.1) holds and the system (2.12) has a solution $d \in \mathbb{R}^n$.*

Proof. Suppose that $\bar{x}$ is PNM regular, that x is sufficiently near $\bar{x}$, and that $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ is an arbitrary partition of $\mathcal{E}^{0+}(x)$. From proposition 3.1, it suffices to prove that MFCQ (3.1) holds at the considered x.

Let $(\mathcal{E}_{\mathcal{F}}(x), \mathcal{E}_{\mathcal{G}}(x))$ be any partition of $\mathcal{E}(x)$ such that (3.8a) holds. Define the partition $(\mathcal{E}_{\mathcal{F}}^{0+}(\bar{x}), \mathcal{E}_{\mathcal{G}}^{0+}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ by the construction (3.8b)–(3.8c), which makes this partition depend on the considered x. By lemma 3.7, the inclusions in (3.10) hold. This implies that the partition $(E_F(x), E_G(x), I(x))$ of $[1:n]$ is a valid instance of partition (E_F, E_G, I) in the definition 3.8 of the PNM regularity. By (3.11), there holds

$$\begin{aligned}&\textstyle\sum_{i\in E_F(x)} \alpha_i \nabla F_i(\bar{x}) + \sum_{i\in E_G(x)} \beta_i \nabla G_i(\bar{x}) + \sum_{i\in I(x)} \big[\alpha_i \nabla F_i(\bar{x}) + \beta_i \nabla G_i(\bar{x})\big] = 0 \\ &\text{and } (\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0.\end{aligned}$$

It is known that this implication is still valid for a small perturbation of the gradients $\nabla F_i(\bar{x})$ and $\nabla G_i(\bar{x})$, resulting from using $\nabla F_i(x)$ and $\nabla G_i(x)$ instead of $\nabla F_i(\bar{x})$ and $\nabla G_i(\bar{x})$ (x is supposed to be close to $\bar{x}$ and F' and G' are continuous at $\bar{x}$). Therefore,

$$\begin{aligned}&\textstyle\sum_{i\in E_F(x)} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(x)} \beta_i \nabla G_i(x) + \sum_{i\in I(x)} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0 \\ &\text{and } (\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0.\end{aligned}$$

This is precisely the MFCQ (3.1). □

Proposition 3.14 below can also be used to get the result of proposition 3.9. Nevertheless, in proposition 3.14, a stronger assumption on F and G is made in order to ensure some radial Lipschitz continuity at $\bar{x}$ of selected solutions to (2.12). Without this stronger assumption, however, the proof of proposition 3.14 still shows the existence of a solution to (2.12) for x near $\bar{x}$, in a constructive manner.

The previous proposition shows that the PNM regularity of a point $\bar{x}$ is a sufficient condition to guarantee a solution to the system (2.12) when x is near $\bar{x}$, but this is certainly not a necessary condition; see counterexample 3.13.

The next counterexample shows that the PNM regularity does not necessarily imply the NM-regularity in the sense of definition 2.1, which guarantees that the Jacobian of the system (2.1), defining a plain NM direction, is nonsingular. As a result, the plain NM direction may not be well defined at x, despite the PNM regularity holds at x.

Counterexample 3.10 (PNM regularity $\not\Rightarrow$ NM-regularity) Consider the LCP (1.2), in which

$$n = 2, \qquad M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \text{and} \qquad q = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Let us make the correspondence between the LCP (1.2) and the general complementarity problem (1.1) by defining F and G at x by $F(x) = Mx + q$ and $G(x) = x$. Then, at $\bar{x} = (-1, -2)$, one has $\mathcal{F}(\bar{x}) = \mathcal{G}(\bar{x}) = \varnothing$ and $\mathcal{E}(\bar{x}) = \{1, 2\}$. Taking $(\tilde{\mathcal{F}}(\bar{x}), \tilde{\mathcal{G}}(\bar{x})) = (\{2\}, \{1\})$ as partition of $\{1, 2\}$ satisfying $\tilde{\mathcal{F}}(\bar{x}) \supseteq \mathcal{F}(\bar{x})$ and $\tilde{\mathcal{G}}(\bar{x}) \supseteq \mathcal{G}(\bar{x})$, the Jacobian of the system (2.1) reads

$$\begin{pmatrix} F_2'(\bar{x}) \\ G_1'(\bar{x}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

This one is singular, showing the $\bar{x}$ is not NM-regular (and in the present case, the system (2.1) has no solution). However, the PNM regularity holds at $\bar{x}$, since $E_F(\bar{x}) = E_G(\bar{x}) = \varnothing$ and $I(\bar{x}) = \{1, 2\}$, so that there is a single acceptable partition (E_F, E_G, I) of $\{1, 2\}$, which is $(\varnothing, \varnothing, \{1, 2\})$. Hence, the premise in (3.11) reads

$$\alpha_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 \qquad \text{and} \qquad (\alpha, \beta) \geqslant 0.$$

This one clearly implies that $\alpha = \beta = 0$, showing the PNM regularity of $\bar{x}$. Interestingly, as expected, the system (2.12) has a solution for $x = \bar{x}$ and for x near $\bar{x}$ (for any x actually) since it consists in the system of inequalities $d_1 \geqslant \max(F_2(x), x_1)$ and $d_2 \geqslant \max(F_1(x), x_2)$, which presents no compatibility problem. □

3.1.2 Regularity of a solution

The definition 3.8 of the PNM regularity has been given for an arbitrary point $\bar{x} \in \mathbb{R}^n$. In this section, we consider the particular case where the considered point $\bar{x}$, whose PNM regularity is examined, is a solution to the complementarity problem (1.1).

The next proposition shows how the concept simplifies when $\bar{x}$ is a solution to the complementarity problem (1.1). Observe that τ no longer intervenes in the property; this is essentially due to the fact that the components of $F(\bar{x})$ and $G(\bar{x})$ are nonnegative, implying that the index set where τ intervenes, namely $I(\bar{x})$, is empty.

Proposition 3.11 (PNM regularity of a solution) *Suppose that $\bar{x} \in \mathbb{R}^n$ is a solution to the complementarity problem (1.1). Then $\bar{x}$ is PNM regular if and only if for any partition (E_F, E_G, I) of $[1\!:\!n]$ such that $\mathcal{F}(\bar{x}) \subseteq E_F$, $\mathcal{G}(\bar{x}) \subseteq E_G$, and $I \subseteq \mathcal{E}^0(\bar{x})$, the implication (3.11) holds.*

Proof. When $\bar{x}$ is a solution to the NCP problem (1.1), one has that $\mathcal{E}_\tau^-(\bar{x}) = \varnothing$ and $(\mathcal{E}_\mathcal{F}^{0+}(\bar{x}), \mathcal{E}_\mathcal{G}^{0+}(\bar{x})) = (\mathcal{E}_\mathcal{F}^{0}(\bar{x}), \mathcal{E}_\mathcal{G}^{0}(\bar{x}))$ is an arbitrary partition of $\mathcal{E}^{0+}(\bar{x}) = \{i : F_i(\bar{x}) = G_i(\bar{x}) = 0\} =: \mathcal{E}^0(\bar{x})$. From the definitions (2.13) and (3.9), we get

$$
\begin{gathered}
I(\bar{x}) = \varnothing, \quad I^\circ(\bar{x}) = \mathcal{E}^0(\bar{x}), \\
E_F(\bar{x}) = \mathcal{F}(\bar{x}) \cup \mathcal{E}_\mathcal{F}^0(\bar{x}), \quad E_F^\circ(\bar{x}) = \mathcal{F}(\bar{x}), \\
E_G(\bar{x}) = \mathcal{G}(\bar{x}) \cup \mathcal{E}_\mathcal{G}^0(\bar{x}), \quad \text{and} \quad E_G^\circ(\bar{x}) = \mathcal{G}(\bar{x}).
\end{gathered}
$$

Then, from definition 3.8, the PNM regularity of $\bar{x}$ reads: for any partition $(\mathcal{E}_\mathcal{F}^0(\bar{x}), \mathcal{E}_\mathcal{G}^0(\bar{x}))$ of $\mathcal{E}^0(\bar{x})$ and any partition (E_F, E_G, I) of $[1\!:\!n]$ such that $\mathcal{F}(\bar{x}) \subseteq E_F \subseteq \mathcal{F}(\bar{x}) \cup \mathcal{E}_\mathcal{F}^0(\bar{x})$, $\mathcal{G}(\bar{x}) \subseteq E_G \subseteq \mathcal{G}(\bar{x}) \cup \mathcal{E}_\mathcal{G}^0(\bar{x})$ and $I \subseteq \mathcal{E}^0(\bar{x})$, the implication (3.11) holds.

The premise of the implication in the proposition gives constraints on the partition (E_F, E_G, I) that look less restrictive than in the previous statement. Let us show that this is not the case, which will conclude the proof.

Suppose indeed that (E_F, E_G, I) is a partition of $[1\!:\!n]$ such that $\mathcal{F}(\bar{x}) \subseteq E_F$, $\mathcal{G}(\bar{x}) \subseteq E_G$, and $I \subseteq \mathcal{E}^0(\bar{x})$, like in the statement of the proposition. If $i \in E_F$, then $i \notin \mathcal{G}(\bar{x})$ [since otherwise it would be in E_G, which is disjoint from E_F to which i belongs], hence i must be in $\mathcal{F}(\bar{x}) \cup \mathcal{E}(\bar{x})$ [since $(\mathcal{F}(\bar{x}), \mathcal{G}(\bar{x}), \mathcal{E}(\bar{x}))$ is a parition of $[1\!:\!n]$], which identical to $\mathcal{F}(\bar{x}) \cup \mathcal{E}^0(\bar{x})$ when $\bar{x}$ is a solution to the complementarity problem. We have shown that $E_F \subseteq \mathcal{F}(\bar{x}) \cup \mathcal{E}^0(\bar{x})$. Similarly $E_G \subseteq \mathcal{G}(\bar{x}) \cup \mathcal{E}^0(\bar{x})$. Now, since $E_F \cap E_G = \varnothing$, one can write that $E_F \subseteq \mathcal{F}(\bar{x}) \cup \mathcal{E}_\mathcal{F}^0(\bar{x})$ and $E_G \subseteq \mathcal{G}(\bar{x}) \cup \mathcal{E}_\mathcal{G}^0(\bar{x})$, where $(\mathcal{E}_\mathcal{F}^0(\bar{x}), \mathcal{E}_\mathcal{G}^0(\bar{x}))$ is a partition of $\mathcal{E}^0(\bar{x})$. Hence the preceding statement is recovered. □

In the very special case where the solution $\bar{x}$ is *nondegenerate*, in the sense that $\mathcal{E}^0(\bar{x}) = \varnothing$ (i.e., there is no index i such that $F_i(\bar{x}) = G_i(\bar{x}) = 0$), the PNM regularity condition of proposition 3.11 contains no choice of partition and simply tells us that

$\sum_{i\in\mathcal{F}(\bar x)} \alpha_i \nabla F_i(\bar x) + \sum_{i\in\mathcal{G}(\bar x)} \beta_i \nabla G_i(\bar x) = 0$ implies that $(\alpha,\beta) = 0$. This is equivalent to the fact that

$$\begin{pmatrix} F'_{\mathcal{F}(\bar x)}(\bar x) \\ G'_{\mathcal{G}(\bar x)}(\bar x) \end{pmatrix} \text{ is nonsingular,}$$

which is a rather weak assumption.

In the other very special case where the solution $\bar x$ is *fully degenerate*, in the sense that $F(\bar x) = G(\bar x) = 0$, there hold $\mathcal{F}(\bar x) = \varnothing$, $\mathcal{G}(\bar x) = \varnothing$, and $\mathcal{E}^0(\bar x) = [1\,{:}\,n]$. Then, the PNM regularity condition of proposition 3.11 becomes:

$$\text{for any partition } (E_F, E_G, I) \text{ of } [1\,{:}\,n], \text{ the implication (3.11) holds.} \tag{3.12}$$

Recall from counterexample 3.10 that the PNM regularity of an arbitrary point does not necessarily imply its NM-regularity in the sense of definition 2.1. According to the next corollary, this implication holds, however, when the considered point is a solution.

Corollary 3.12 (at a solution, PNM-regularity $\Rightarrow$ NM-regularity) *Suppose that $\bar x \in \mathbb{R}^n$ is a PNM regular solution to the complementarity problem (1.1). Then, $\bar x$ is NM-regular.*

Proof. By taking $I = \varnothing$ in the expression of the PNM regularity at a solution $\bar x$ given by proposition 3.11, we see that for any partition (E_F, E_G) of $[1\,{:}\,n]$ such that $\mathcal{F}(\bar x) \subseteq E_F$ and $\mathcal{G}(\bar x) \subseteq E_G$, we have that any (α,β) verifying $\sum_{i\in E_F} \alpha_i \nabla F_i(\bar x) + \sum_{i\in E_G} \beta_i \nabla G_i(\bar x) = 0$ vanish. This is equivalent to saying that the matrix

$$\begin{pmatrix} F'_{E_F}(\bar x) \\ G'_{E_G}(\bar x) \end{pmatrix}$$

is nonsingular. In view of the possible choices of the index sets E_F and E_G, we see that this is exactly the NM-regularity of definition 2.1 (the partition (E_F, E_G) of $[1\,{:}\,n]$ plays the role of $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ in the definition). $\square$

The goal of the next counterexample is twofold. Firstly, it shows that the NM-regularity of a solution does not imply its PNM regularity; hence, the reciprocal of the implication in corollary 3.12 does not hold in general. Secondly, it shows that the PNM regularity of a solution $\bar x$ is not a necessary condition for guaranteeing that the system (2.12) has a solution d for x near $\bar x$; hence, according to proposition 3.9, the PNM regularity is just a sufficient condition for the solvability of (2.12).

Counterexample 3.13 (consistency of (2.12) without PNM regularity) Consider the trivial LCP (1.2) with $n = 1$, $M < 0$ and $q = 0$. Then $\bar x = 0$ is the unique solution to that problem. We have the following properties.

1. *The point $\bar x = 0$ is NM-regular.* Indeed $\mathcal{F}(\bar x) = \mathcal{G}(\bar x) = \varnothing$ and $\mathcal{E}(\bar x) = \mathcal{E}^0(\bar x) = \{1\}$. The claim now follows from the observation that the two possible Jacobians in (2.2), namely 1 and M, are nonsingular.

2. *The point $\bar{x} = 0$ is not PNM regular.* Indeed, taking $E_F = E_G = \varnothing$ and $I = \mathcal{E}^0(\bar{x}) = \{1\}$ in the expression of the PNM regualrity of the solution $\bar{x}$, given by proposition 3.11, we should have (3.11), that is

$$\alpha_1 + \beta_1 M = 0, \quad \alpha_1 \geqslant 0, \quad \beta_1 \geqslant 0 \qquad \Longrightarrow \qquad \alpha_1 = \beta_1 = 0,$$

which is not true with $M < 0$ (actually, it will be shown with proposition **??** that a fully degenerate solution of an LCP, like $\bar{x}$ here, is PNM regular if and only if M is a **P**-matrix, which is not the case here).

3. *The system* (2.12) *has a solution for any* x. Indeed, for the given problem, the system (2.12) has one of the following forms, depending on the index sets $E_F(x)$, $E_G(x)$, and $I(x)$:

$$\begin{aligned} x + d &= 0, && \text{if } 1 \in E_F(x), \\ Mx + Md &= 0, && \text{if } 1 \in E_G(x), \\ x + d \geqslant 0 \quad \text{and} \quad Mx + Md &\geqslant 0, && \text{if } 1 \in I(x). \end{aligned}$$

Each of these systems has for unique solution $d = -x$ (hence, $x + d$ is the unique solution to the considered LCP), showing that the system (2.12) is always consistent, despite the fact that $\bar{x}$ is not PNM regular. $\square$

3.1.3 Radial Lipschitz continuity of the directions

We now consider the question to know whether a solution d to (2.12) at x can be chosen in such a way that it remains bounded when x is in a neighborhood of a given point $\bar{x}$. The next proposition shows that this property is verified if the regularity condition (3.11) holds at $\bar{x}$. It is actually shown that, one can *find* a solution to (2.12) at a point x near $\bar{x}$, such that d is near some solution $\bar{d}$ to (2.12) at $x = \bar{x}$. It is therefore a kind of continuity property. The proof of the proposition is a refinement of the one of proposition 3.9.

We shall say that a function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is *radially Lipschitz continuous* at $\bar{x} \in \mathbb{R}^n$ if there is a neighborhood V of $\bar{x}$ in $\mathbb{R}^n$ and a constant $L \geqslant 0$, such that for all $x \in V$, $\|\varphi(x) - \varphi(\bar{x})\| \leqslant L\|x - \bar{x}\|$.

Proposition 3.14 (radial Lipschitz continuity of the directions) *Suppose that F and G are differentiable near some $\bar{x} \in \mathbb{R}^n$, that F' and G' are continuous at $\bar{x}$, and that the regularity condition* (3.11) *holds at $\bar{x}$. Then, the following properties hold.*

(1) *For any partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ and any partition (E_F, E_G, I) of $[1\,{:}\,n]$ such that $E^\circ_F(\bar{x}) \subseteq E_F \subseteq E_F(\bar{x})$, $E^\circ_G(\bar{x}) \subseteq E_G \subseteq E_G(\bar{x})$, and $I(\bar{x}) \subseteq I \subseteq I^\circ(\bar{x})$, the system*

$$\begin{cases} F_i(\bar{x}) + F'_i(\bar{x})\bar{d} = 0 & \text{if } i \in E_F \\ G_i(\bar{x}) + G'_i(\bar{x})\bar{d} = 0 & \text{if } i \in E_G \\ F_i(\bar{x}) + F'_i(\bar{x})\bar{d} > 0 & \text{if } i \in I \\ G_i(\bar{x}) + G'_i(\bar{x})\bar{d} > 0 & \text{if } i \in I \end{cases} \tag{3.13}$$

has a solution $\bar{d}$. Denote by $\bar{D}$ the finite set of these $\bar{d}$'s, each of them corresponding to one of the partitions given above.

(2) *For any $\delta > 0$, there is a neighborhood V of $\bar{x}$ such that, for any $x \in V$, the system (2.12) has a solution d that satisfies*

$$\min_{\bar{d}\in\bar{D}} \|d - \bar{d}\| < \delta.$$

(3) *If, in addition, F' and G' are radially Lipschitz continuous at $\bar{x}$, then, there is a neighborhood V' of $\bar{x}$ and a constant $L \geqslant 0$ such that, for any $x \in V'$, the system (2.12) has a solution d that satisfies*

$$\min_{\bar{d}\in\bar{D}} \|d - \bar{d}\| \leqslant L\|x - \bar{x}\|.$$

Proof. Let $\tau \in (0, \infty]$ and denote by V_0 the neighborhood of $\bar{x}$ on which F and G are differentiable (and their derivatives are radially Lipschitz continuous at $\bar{x}$ in case (3)).

[(1)] For an arbitrary partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$, define the partition $(E_F(\bar{x}), E_G(\bar{x}), I(\bar{x}))$ of $[1\!:\!n]$ by (2.13) at $x = \bar{x}$. Define also the partition $(E^\circ_F(\bar{x}), E^\circ_G(\bar{x}), I^\circ(\bar{x}))$ of $[1\!:\!n]$ by (3.9). By the regularity condition (3.11), for any partition (E_F, E_G, I) of $[1\!:\!n]$ such that $E^\circ_F(\bar{x}) \subseteq E_F \subseteq E_F(\bar{x})$, $E^\circ_G(\bar{x}) \subseteq E_G \subseteq E_G(\bar{x})$, and $I(\bar{x}) \subseteq I \subseteq I^\circ(\bar{x})$, the linear system (3.13) has a solution. Choose one of these solutions and denote it by $\bar{d}$ (there are actually an unbounded set of solutions to the above system when $I \neq \varnothing$). Now, denote by $\bar{D}$ the set formed of all the selected solutions $\bar{d}$, each of them corresponding to one partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ and one partition (E_F, E_G, I) of $[1\!:\!n]$ specified above. The set $\bar{D}$ is finite, because there is a finite number of such partitions.

[(2)] For any partition (E_F, E_G, I) of $[1\!:\!n]$ determined like in point (1) and any $x \in V_0$, denote by $\mathcal{S}(x)$ the (possibly empty) affine set of solutions d to the linear equations

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in E_F \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in E_G. \end{cases} \tag{3.14}$$

To make the notation compact, let us write this linear system in d by $A(x)d = b(x)$, with

$$A(x) := \begin{pmatrix} F'_{E_F}(x) \\ G'_{E_G}(x) \end{pmatrix} \quad \text{and} \quad b(x) := -\begin{pmatrix} F_{E_F}(x) \\ G_{E_G}(x) \end{pmatrix}.$$

The regularity (3.11) at $\bar{x}$ implies that $A(\bar{x})$ is surjective (see the equivalence between (3.1) and (3.2)), so that, for x in a neighborhood $V_1 \subseteq V_0$ of $\bar{x}$, $A(x)$ is also surjective, implying that $\mathcal{S}(x)$ is nonempty.

For x in a neighborhood $V_2 \subseteq V_1$ of $\bar{x}$, lemma 3.4 allows us to construct a partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ such that (3.8b) and (3.8c) holds and, by lemma 3.7, such that the inclusions in (3.10) hold. This implies that the partition $(E_F(x), E_G(x), I(x))$ of $[1\!:\!n]$ is a valid instance of partition (E_F, E_G, I) determined like in point (1). Let $\bar{d} \in \bar{D}$ be the solution to (3.13) that was associated with these partitions in point (1). One can give an explicit expression of the projection of $\bar{d}$ onto $\mathcal{S}(x)$:

$$d := \mathrm{P}_{\mathcal{S}(x)}\, \bar{d} = \mathrm{P}_{\mathcal{N}(A(x))}\, \bar{d} + b^\circ(x), \tag{3.15}$$

where $\mathrm{P}_{\mathcal{A}}$ denotes the affine or linear orthogonal projector on a nonempty affine or linear space $\mathcal{A} \subseteq \mathbb{R}^n$, $\mathcal{N}(A(x))$ denotes the null space of $A(x)$, and $b^\circ(x) := A(x)^\mathsf{T} \left(A(x)A(x)^\mathsf{T}\right)^{-1} b(x)$. Since $\bar{d} = \mathrm{P}_{\mathcal{S}(\bar{x})}\,\bar{d}$, it follows that

$$\|d - \bar{d}\| \leqslant \|\,\mathrm{P}_{\mathcal{N}(A(x))} - \mathrm{P}_{\mathcal{N}(A(\bar{x}))}\,\|\,\|\bar{d}\| + \|b^\circ(x) - b^\circ(\bar{x})\|. \tag{3.16}$$

By the continuity of F, G, F', and G' at $\bar{x}$, the orthogonal projector on $\mathcal{N}(A(x))$ and b° are continuous at $\bar{x}$, so that $\mathrm{P}_{\mathcal{S}(x)}$ is also continuous at $\bar{x}$. Therefore, for any $\delta > 0$, there exists a neighborhood $V_3 \subseteq V_2$ of $\bar{x}$ such that $\|d - \bar{d}\| < \delta$. This neighborhood V_3 depends on the partition $(E_F(x), E_G(x), I(x))$ of $[1\,{:}\,n]$ associated with x, hence on the partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ associated with it. Since the number of such latter partitions is finite, the last claim remains valid (one takes the intersection V of all the possible neighborhoods V_3's).

[(3)] The reasoning is identical to the one presented in point (2). But now, one can use the radial Lipschitz property of F, G, F', and G' at $\bar{x}$ to deduce from (3.15) the existence of a neighborhood $V_4 \subseteq V_2$ and a constant $L \geqslant 0$, such that, for $x \in V_4$:

$$\|d - \bar{d}\| \leqslant L\,\|x - \bar{x}\|.$$

Like in point (2), the neighborhood V_4 and the constant L depend on the partition $(E_F(x), E_G(x), I(x))$ of $[1\,{:}\,n]$ associated with x, hence on the partition $(\mathcal{E}^{0+}_{\mathcal{F}}(\bar{x}), \mathcal{E}^{0+}_{\mathcal{G}}(\bar{x}))$ of $\mathcal{E}^{0+}(\bar{x})$ associated with it. Since the number of such latter partitions is finite, the last claim remains valid (one takes the intersection V' of all the possible neighborhoods V_4's and the maximum of the constants L's). □

The next property will be useful for establishing the global convergence result of theorems 3.17 and 3.18.

Proposition 3.15 (local boundedness of the directions) *Suppose that F and G are differentiable near some $\bar{x} \in \mathbb{R}^n$, that F' and G' are continuous at $\bar{x}$, and that the regularity condition (3.11) holds at $\bar{x}$. Then, there is a constant C, such that, for x near $\bar{x}$, the system (2.12) has a solution d that satisfies $\|d\| \leqslant C$.*

Proof. It is a consequence of point (2) of proposition 3.14, since $\bar{D}$ is bounded by its finite number of elements. □

3.2 Global convergence

The global convergence results of this section accept directions d such that the right-hand side of (2.18) is sufficiently negative in the sense of (2.19a), an inequality that we reproduce here for the reader's convenience:

$$-\sum_{i\in[1\,:\,n]} (1 - \rho_i(x,d))H_i(x)^2 \leqslant -2(1-\eta)\,\theta(x), \tag{3.17}$$

where $\rho_i(x,d)$ is defined by (2.20), H is the function defined by (1.3b), and η is a constant (independent of k) such that $\eta < 1$. By proposition 2.18, this inequality implies that d is a descent direction of θ at x, since then

$$\theta'(x;d) \leqslant -2(1-\eta)\,\theta(x), \tag{3.18}$$

and the right-hand side is negative when $\theta(x) \neq 0$, that is when x is not a solution to the NCP (1.1). It would have been less restrictive to impose the satisfaction of (3.18), instead of that of (3.17), but the technique used in the proof of theorem 3.16 below would have then required to have a reverse inequality in (2.18) in order to recover (3.17), which is the inequality that is required in the adopted proof; the reverse inequality in (2.18) looked problematic to us. Note that the inequality (3.17) simplifies into (2.19b).

We start the global convergence analysis with theorem 3.16, which assumes that the generic algorithm 2.11 generates a sequence $\{x_k\}$, hence is well-posed, and the boundedness of the direction subsequence $\{d_k\}_{k\in\mathcal{K}}$ when the subsequence $\{x_k\}_{k\in\mathcal{K}}$ of $\{x_k\}$ converges to some point $\bar{x}$. The convergence of the algorithms 2.12 and 2.13 will be examined in theorems 3.17 and 3.18, respectively. The proof of theorem 3.16 contains the main arguments. We have preferred presenting the convergence result in two stages (theorem 3.16 and theorems 3.17 and 3.18), since the boundedness assumption may be due to the structure of the problem, making the lemma useful in that circumstance. In theorems 3.17 and 3.18, which can also be viewed as corollaries of theorem 3.16, it is the assumed regularity of the limit point $\bar{x}$ that ensures the boundedness of $\{d_k\}_{k\in\mathcal{K}}$ and therefore the global convergence of the algorithm. These global convergence results of theorems 3.17 and 3.18 are rather weak since they assume that the generated sequence has a limit point (this will be certainly the case when this sequence is bounded) and that the limit point is regular in a certain sense. It may occur, however, that the generated sequence $\{x_k\}$ has no regular limit points, in which case the theorem provides no information. Nevertheless, it acts as a filter that the algorithms must pass, which was very useful to us in the design of an acceptation test (2.19)-(3.17) for the hybrid algorithm 2.13.

Theorem 3.16 (global convergence of the generic algorithm) *Let F and $G : \Omega \to \mathbb{R}^n$ be differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the generic algorithm 2.11 generates a sequence $\{x_k\} \subseteq \Omega$. If $\bar{x} \in \Omega$ is an accumulation point of $\{x_k\}$, at which F' and G' are continuous, and if the subsequence $\{d_k : x_k$ is near $\bar{x}\}$ is bounded, then all the sequence $\{\theta(x_k)\}_{k\geqslant 1}$ converges to zero and $\bar{x}$ is a solution to (1.1).*

Proof. By the Armijo inequality (2.23), the sequence $\{\theta(x_k)\}$ decreases; since this sequence is also bounded below (by zero), it converges. By the Armijo inequality (2.23) again and the fact that $\eta < 1$, it follows that

$$\lim_{k\to\infty} \alpha_k\,\theta(x_k) = 0. \tag{3.19}$$

Let us examine two complementary cases.

If $\limsup_{k\to\infty} \alpha_k > 0$ (or, equivalently, $\alpha_k \not\to 0$), there is a subsequence $\mathcal{K}' \subseteq \mathbb{N}$ such that $\{\alpha_k\}_{k\in\mathcal{K}'}$ is bounded away from zero. Then, (3.19) implies that $\lim_{k\to\infty,\,k\in\mathcal{K}'} \theta(x_k) = 0$

and actually $\lim_{k\to\infty} \theta(x_k) = 0$, since the sequence $\{\theta(x_k)\}$ decreases. By the continuity of θ, any accumulation point $\bar{x}$ of $\{x_k\}$ satisfies $\theta(\bar{x}) = 0$, which means that $\bar{x}$ solves (1.1). We have shown the conclusions of the theorem in that case.

We now consider the more difficult case when $\limsup_{k\to\infty} \alpha_k = 0$ (or, equivalently, $\alpha_k \to 0$). Let us first sketch the proof, which is inspired from that in [45]; see also [69]. Let $\{x_k\}_{k\in\mathcal{K}}$ be a subsequence converging to $\bar{x}$ ($k \to \infty$ in some infinite subset $\mathcal{K}$ of $\mathbb{N}$) With no loss of generality, one can assume that $\alpha_k < 1$, which implies that the stepsize $\hat{\alpha}_k := \alpha_k/\beta$ is rejected by the Armijo rule. Of course $\hat{\alpha}_k \to 0$. Let $\hat{x}_k := x_k + \hat{\alpha}_k d_k$ be the corresponding rejected point. Then, $\theta(\hat{x}_k) > \theta(x_k) - 2\omega\hat{\alpha}_k(1-\eta)\,\theta(x_k)$ or

$$4\omega\hat{\alpha}_k(1-\eta)\,\theta(x_k) > 2[\theta(x_k) - \theta(\hat{x}_k)]. \tag{3.20}$$

The tactic of the proof consists in writing the right-hand side of this inequality as follows

$$2[\theta(x_k) - \theta(\hat{x}_k)] = \sum_{i=1}^{n} \left[\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2\right] \tag{3.21}$$

and to find a lower bound of each term of the sum in the right-hand side of the previous identity. More specifically, we shall show that, since $\{d_k\}_{k\in\mathcal{K}}$ is assumed to be bounded, for any $i \in [1\!:\!n]$ and any iterate x_k sufficiently close to $\bar{x}$, the following inequality holds

$$\begin{aligned} &\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\ &\qquad \geqslant\ 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k), \end{aligned} \tag{3.22}$$

where $\rho_{k,i} := \rho_i(x_k, d_k)$ and the term $o(\hat{\alpha}_k)$ means that $o(\hat{\alpha}_k)/\hat{\alpha}_k \to 0$ when $k \to \infty$ in $\mathcal{K}$. Then, the inequality (3.20), with its right-hand side bounded below thanks to the identity (3.21) and the inequalities (3.22), yields

$$\begin{aligned} &4\omega\hat{\alpha}_k(1-\eta)\,\theta(x_k) \\ &\quad \geqslant 2\hat{\alpha}_k \sum_{i\in[1\,:\,n]} (1-\rho_{k,i}) \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [(3.20),\ (3.21),\ (3.22)] \\ &\quad \geqslant 4\hat{\alpha}_k(1-\eta)\,\theta(x_k) + o(\hat{\alpha}_k). \qquad [(2.19)] \end{aligned}$$

After division by $4\hat{\alpha}_k(1-\eta)$, we get

$$\omega\,\theta(x_k) \geqslant \theta(x_k) + \frac{o(\hat{\alpha}_k)}{\hat{\alpha}_k}. \tag{3.23}$$

Taking the limit when $k \to \infty$ in $\mathcal{K}$ shows that $\omega\,\theta(\bar{x}) \geqslant \theta(\bar{x})$. Since $\omega \in (0,1)$ and $\theta(\bar{x}) \geqslant 0$, this implies that $\theta(\bar{x}) = 0$. Therefore, all the sequence $\{\theta(x_k)\}$ tends to zero and $\bar{x}$ solves (1.1). We have also shown the conclusions of the theorem in that case.

Therefore, to conclude the proof, we only have to show (3.22), *for all* $i \in [1\!:\!n]$ *and* x_k *sufficiently close to* $\bar{x}$.

Since $\{d_k\}_{k\in\mathcal{K}}$ is bounded by assumption and $\alpha_k \to 0$, it follows that $\hat{x}_k \to \bar{x}$ when $k \to \infty$ in $\mathcal{K}$. Now, for $i \in [1\!:\!n]$, the mean value theorem provides

$$|F_i(\hat{x}_k) - F_i(x_k) - F_i'(x_k)(\hat{x}_k - x_k)| \leqslant \left(\sup_{z\in(x_k,\hat{x}_k)} \|F_i'(z) - F_i'(x_k)\|\right) \|\hat{x}_k - x_k\|.$$

A similar estimation holds for G_i. By the continuity of F' at $\bar{x}$, the factor in parenthesis in the right-hand side tends to zero when $k \to \infty$ in $\mathcal{K}$. Using $\hat{x}_k - x_k = \hat{\alpha}_k d_k$, we get

$$F_i(\hat{x}_k) = F_i(x_k) + \hat{\alpha}_k F_i'(x_k) d_k + o(\hat{\alpha}_k),$$
$$G_i(\hat{x}_k) = G_i(x_k) + \hat{\alpha}_k G_i'(x_k) d_k + o(\hat{\alpha}_k).$$

Below, we shall need to give a lower bound on $F_i(x_k)^2 - F_i(\hat{x}_k)^2$ and $G_i(x_k)^2 - G_i(\hat{x}_k)^2$. By the previous estimates, we have

$$F_i(x_k)^2 - F_i(\hat{x}_k)^2 = -2\hat{\alpha}_k F_i(x_k) F_i'(x_k) d_k + o(\hat{\alpha}_k), \tag{3.24a}$$
$$G_i(x_k)^2 - G_i(\hat{x}_k)^2 = -2\hat{\alpha}_k G_i(x_k) G_i'(x_k) d_k + o(\hat{\alpha}_k). \tag{3.24b}$$

Let us now examine each term of the sum in (3.21) for the indices i in the following partition of $[1:n]$:

$$\Big(\mathcal{F}(\bar{x}),\ \ \mathcal{G}(\bar{x}),\ \ \mathcal{E}^+(\bar{x}),\ \ \mathcal{E}^-(\bar{x}),\ \ \mathcal{E}^0(\bar{x})\Big).$$

Note that τ does not intervene in this partition.

1. $i \in \mathcal{F}(\bar{x})$.

   By the strict inequality $F_i(\bar{x}) < G_i(\bar{x})$ defining $\mathcal{F}(\bar{x})$ in (1.8), the continuity of F and G at $\bar{x}$, and the fact that x_k is close to $\bar{x}$ when k is large in $\mathcal{K}$, we have $F_i(x_k) < G_i(x_k)$ or $i \in \mathcal{F}(x_k)$ for large k in $\mathcal{K}$. Let us show that

   $$-\,F_i(x_k) F_i'(x_k) d_k \geqslant (1-\rho_{k,i}) F_i(x_k)^2. \tag{3.25}$$

   We examine three complementary cases.

   - If $F_i(x_k) = 0$, (3.25) is clearly verified with equality.
   - If $i \in \mathcal{F}(x_k) \setminus \mathcal{E}_\tau^-(x_k)$ and $F_i(x_k) \neq 0$, $(2.20)_1$ gives $F_i'(x_k) d_k = -(1-\rho_{k,i}) F_i(x_k)$. Multiplying both sides of this equality by $-F_i(x_k)$ yields (3.25) with equality.
   - If $i \in \mathcal{F}^-(x_k) \cap \mathcal{E}_\tau^-(x_k)$ (in which case $F_i(x_k) < 0$), $(2.20)_5$ gives $F_i'(x_k) d_k \geqslant -(1-\rho_{k,i}) F_i(x_k)$. Multiplying both sides of this inequality by $-F_i(x_k) > 0$ yields (3.25).

   Since $\hat{x}_k \to \bar{x}$ when $k \to \infty$ in $\mathcal{K}$ and since $F_i(\bar{x}) < G_i(\bar{x})$ when $i \in \mathcal{F}(\bar{x})$, one also has $F_i(\hat{x}_k) < G_i(\hat{x}_k)$. Therefore,

   $$\begin{aligned}
   &\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\
   &\quad = \ F_i(x_k)^2 - F_i(\hat{x}_k)^2 \qquad [F_i(x_k) < G_i(x_k) \text{ and } F_i(\hat{x}_k) < G_i(\hat{x}_k)] \\
   &\quad = \ -2\hat{\alpha}_k F_i(x_k) F_i'(x_k) d_k + o(\hat{\alpha}_k) \qquad [(3.24a)] \\
   &\quad \geqslant \ 2(1-\rho_{k,i}) \hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.25)] \\
   &\quad = \ 2(1-\rho_{k,i}) \hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) < G_i(x_k)].
   \end{aligned}$$

   We have obtained the desired inequality (3.22).

2. $i \in \mathcal{G}(\bar{x})$.

   One can proceed like in case 1, by switching the roles of F and G.

3. $i \in \mathcal{E}^+(\bar{x})$.

In this case, $F_i(x_k)$, $G_i(x_k)$, $F_i(\hat{x}_k)$, and $G_i(\hat{x}_k)$ are positive for k large in $\mathcal{K}$, which implies that i is in one of the sets $\mathcal{F}^+(x_k) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x_k)$ or $\mathcal{G}^+(x_k) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x_k)$. We now consider these sets one after the other.

3.1. $i \in \mathcal{F}^+(x_k) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x_k)$.

In this case, $0 < F_i(x_k) \leqslant G_i(x_k)$. Because $i \in [\mathcal{F}(x_k) \setminus \mathcal{E}^-_{\tau_k}(x_k)] \cup \mathcal{E}^{0+}_{\mathcal{F}}(x_k)$ and $F_i(x_k) \neq 0$, $(2.20)_1$ tells us that $F_i'(x_k)d_k = -(1-\rho_{k,i})F_i(x_k)$ and finally

$$-F_i(x_k)F_i'(x_k)d_k = (1-\rho_{k,i})F_i(x_k)^2. \tag{3.26}$$

Therefore, for k large in $\mathcal{K}$:

$$\begin{aligned}
&\underbrace{\min(F_i(x_k), G_i(x_k))}_{=F_i(x_k)}{}^2 - \underbrace{\min(F_i(\hat{x}_k), G_i(\hat{x}_k))}_{0\leqslant\,\cdot\,\leqslant F_i(\hat{x}_k)}{}^2 \\
&\quad\geqslant\quad F_i(x_k)^2 - F_i(\hat{x}_k)^2 \\
&\quad=\quad -2\hat{\alpha}_k F_i(x_k)F_i'(x_k)d_k + o(\hat{\alpha}_k) \qquad [(3.24a)] \\
&\quad=\quad 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.26)] \\
&\quad=\quad 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) \leqslant G_i(x_k)].
\end{aligned}$$

We have obtained the desired inequality (3.22).

3.2. $i \in \mathcal{G}^+(x_k) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x_k)$.

One can proceed like in case 3.1, by switching the roles of F and G.

4. $i \in \mathcal{E}^-(\bar{x})$.

In this case, for k large in $\mathcal{K}$, $F_i(x_k)$, $G_i(x_k)$, $F_i(\hat{x}_k)$, and $G_i(\hat{x}_k)$ are negative, and $|F_i(x_k)-G_i(x_k)| < \tau$, so that $i \in \mathcal{E}^-_\tau(x_k)$. Then, by $(2.20)_5$, $F_i'(x_k)d_k \geqslant -(1-\rho_{k,i})F_i(x_k)$ and $G_i'(x_k)d_k \geqslant -(1-\rho_{k,i})G_i(x_k)$, so that

$$-F_i(x_k)F_i'(x_k)d_k \geqslant (1-\rho_{k,i})F_i(x_k)^2, \tag{3.27a}$$

$$-G_i(x_k)G_i'(x_k)d_k \geqslant (1-\rho_{k,i})G_i(x_k)^2. \tag{3.27b}$$

Now, one (or both) of the following two cases must occur.

4.1. $F_i(x_k) \leqslant G_i(x_k)$, which are both negative. We divide the analysis of this case into two complementary subcases.

4.1.1. $F_i(\hat{x}_k) \leqslant G_i(\hat{x}_k)$, which are both negative.

For k large in $\mathcal{K}$, the following holds

$$\begin{aligned}
&\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\
&\quad=\quad F_i(x_k)^2 - F_i(\hat{x}_k)^2 \qquad [F_i(x_k) \leqslant G_i(x_k) \text{ and } F_i(\hat{x}_k) \leqslant G_i(\hat{x}_k)] \\
&\quad=\quad -2\hat{\alpha}_k F_i(x_k)F_i'(x_k)d_k + o(\hat{\alpha}_k) \qquad [(3.24a)] \\
&\quad\geqslant\quad 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.27a)] \\
&\quad=\quad 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) \leqslant G_i(x_k)].
\end{aligned}$$

We have obtained the desired inequality (3.22).

4.1.2. $G_i(\hat{x}_k) < F_i(\hat{x}_k)$, which are both negative.

Let us show that

$$2(1-\rho_{k,i})\hat{\alpha}_k \leqslant 1, \quad \text{for } k \text{ large in } \mathcal{K}. \tag{3.28}$$

This is certainly the case when $\rho_{k,i} \geqslant 0$, since then, $2(1-\rho_{k,i})\hat{\alpha}_k \leqslant 2\hat{\alpha}_k \leqslant 1$ because $\hat{\alpha}_k \to 0$ for $k \to \infty$ in $\mathcal{K}$. When $\rho_{k,i} < 0$, we use the fact that the $\rho_{k,i}$ used in this case verifies (see in the preamble of point 4):

$$\rho_{k,i}\, F_i(x_k) \leqslant F_i(x_k) + F_i'(x_k)d_k.$$

Hence, for k large enough in $\mathcal{K}$:

$$\frac{1}{2}\,\rho_{k,i}\, F_i(\bar{x}) \leqslant \rho_{k,i}\, F_i(x_k) \leqslant F_i(x_k) + F_i'(x_k)d_k \leqslant C$$

where the constant $C > 0$ comes for the fact that $x_k \to \bar{x}$ for $k \to \infty$ in $\mathcal{K}$, from the assumed continuity of F' at $\bar{x}$, and from the assumed boundedness of $\{d_k\}$. This shows that $\rho_{k,i}$ is bounded below, so that (3.28) also holds when $\rho_{k,i} < 0$. Then, the following holds

$$\begin{array}{ll}
\multicolumn{2}{l}{\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2} \\
= & F_i(x_k)^2 - G_i(\hat{x}_k)^2 \qquad [F_i(x_k) \leqslant G_i(x_k) \text{ and } G_i(\hat{x}_k) < F_i(\hat{x}_k)] \\
= & G_i(x_k)^2 - G_i(\hat{x}_k)^2 + F_i(x_k)^2 - G_i(x_k)^2 \\
= & -2\hat{\alpha}_k G_i(x_k) G_i'(x_k) d_k + F_i(x_k)^2 - G_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.24b)] \\
\geqslant & 2(1-\rho_{k,i})\hat{\alpha}_k G_i(x_k)^2 + F_i(x_k)^2 - G_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.27b)] \\
= & 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + (1 - 2(1-\rho_{k,i})\hat{\alpha}_k)(F_i(x_k)^2 - G_i(x_k)^2) + o(\hat{\alpha}_k) \\
\geqslant & 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.28) \text{ and } F_i(x_k)^2 \geqslant G_i(x_k)^2] \\
= & 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) \leqslant G_i(x_k)].
\end{array}$$

We have obtained the desired inequality (3.22).

4.2. $G_i(x_k) \leqslant F_i(x_k)$, which are both negative. One can proceed like in case 4.1, by switching the roles of F and G.

5. $i \in \mathcal{E}^0(\bar{x})$.

In this case, we write

$$\begin{aligned}
&\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\
&\quad = \Big(\min(F_i(x_k), G_i(x_k)) - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))\Big) \\
&\qquad \times \Big(\min(F_i(x_k), G_i(x_k)) + \min(F_i(\hat{x}_k), G_i(\hat{x}_k))\Big).
\end{aligned}$$

Since $x \mapsto \min(F(x), G(x))$ is Lipschitz continuous near $\bar{x}$, the first factor in the right-hand side is bounded by a constant times $\|\hat{x}_k - x_k\|$, which is an $O(\hat{\alpha}_k)$ by the boundedness of $\{d_k\}$, while the second factor in the right-hand side converges to zero (since in this case $F_i(\bar{x}) = G_i(\bar{x}) = 0$). Thus the whole term is $o(\hat{\alpha}_k)$. This is enough to get (3.22), since the first term in the right-hand side of (3.22) is also $o(\hat{\alpha}_k)$. □

Theorem 3.17 (global convergence of the PNM algorithm) *Let F and G : $\Omega \to \mathbb{R}^n$ be differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the PNM algorithm 2.12 generates a sequence $\{x_k\} \subseteq \Omega$. If $\bar{x} \in \Omega$ is an accumulation point of $\{x_k\}$ that is PNM regular in the sense of definition 3.8, and if F' and G' are continuous at $\bar{x}$, then all the sequence $\{\theta(x_k)\}_{k\geqslant 1}$ converges to zero and $\bar{x}$ is a solution to (1.1).*

PROOF. According to theorem 3.16, we just have to prove that the subsequence $\{d_k : x_k$ is near $\bar{x}\}$ is bounded. Since the directions d_k are computed by (2.24), this property is given by proposition 3.15, which rests on the additional assumption on the regularity of $\bar{x}$ in the sense (3.11). □

Theorem 3.18 (global convergence of the HNM algorithm) *Let F and G : $\Omega \to \mathbb{R}^n$ be differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the HNM algorithm 2.13 generates a sequence $\{x_k\} \subseteq \Omega$. If $\bar{x} \in \Omega$ is an accumulation point of $\{x_k\}$ that is NM and PNM regular in the sense of definitions 2.1 and 3.8, and if F' and G' are continuous at $\bar{x}$, then all the sequence $\{\theta(x_k)\}_{k\geqslant 1}$ converges to zero and $\bar{x}$ is a solution to (1.1).*

PROOF. According to theorem 3.16, we just have to prove that the subsequence $\{d_k : x_k$ is near $\bar{x}\}$ is bounded. Recall that, in the HNM algorithm 2.13, the direction is computed either as the solution to the linear system (2.1) or as the solution to the optimization problem (2.24).

When d_k is the solution to the system (2.1), the boundedness property of d_k is given by proposition 2.3. When d_k is the solution to problem (2.24), the boundedness property of d_k is given by proposition 3.15 (like in the proof of theorem 3.17). □

Remark 3.19 Let us stress the fact that the previous theorem 3.16 and theorems 3.17 and 3.18 assume that the algorithm generates a sequence $\{x_k\} \subseteq \Omega$, which implicitely supposes that a direction d_k can be computed at each iteration. For a linear complementarity problem of the form (1.2), this assumption is guaranteed for the plain Newton-min direction (2.1), when M is nondegenerate, but this assumption on M may not be sufficient for being able to compute an inexact direction (2.22). Assume indeed that $n = 1$, $M = -1$ (nondegeneracy but not **P**-matricity), and $q = -1$. Then, problem (1.2) has no solution. Consider the point $\bar{x} = -1/2$, which is the minimizer of the merit function θ defined by (1.4). One has $1 \in \mathcal{E}_\tau^-(\bar{x})$, so that the direction in (2.22) must satisfy

$$0 \leqslant (1 - \rho_{k,i})(-1/2) + d \qquad \text{and} \qquad 0 \leqslant (1 - \rho_{k,i})(-1/2) - d,$$

which forms an incompatible system. Therefore, theorem 3.16 and theorem 3.18 do not apply to that problem instance. □

4 Conclusion

This paper presents globally convergent algorithms for solving the complementarity problem (1.1), based on semismooth-like iterations on the nonsmooth equation (1.3), reformulating the problem with the minimum function. In practice, this solution strategy is often more efficient than with other reformulations but it is difficult to implement up to completeness, because the associated least-square merit function may not decrease along the semismooth direction. The paper proposes to overcome the difficulty by slightly modifying this direction in the neighborhood of the kinks of the minimum function. A global convergence result can be established, provided some specific regularity condition holds at the accumulation points of the generated sequence. The algorithms can also be used to solve linear complementarity problems.

A number of issues still need to be considered to improve the robustness of the proposed algorithms, to finalize their analysis, to estimate their complexity, and to highlight their attractiveness. Some of them are explored in [78, 32, 34, 33], and others will be considered in subsequent contributions.

References

[1] V. Acary, B. Brogliato (2008). *Numerical Methods for Nonsmooth Dynamical Systems - Applications in Mechanics and Electronics*. Lecture Notes in Applied and Computational Mechanics 35. Springer. [doi]. 4

[2] M. Aganagić (1984). Newton's method for linear complementarity problems. *Mathematical Programming*, 28, 349–362. [doi]. 4

[3] L. Armijo (1966). Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16, 1–3. [journal]. 21

[4] L. Beaude, K. Brenner, S. Lopez, R. Masson, F. Smai (2018). Non-isothermal compositional liquid gas Darcy flow: formulation, soil-atmosphere boundary condition and application to high-energy geothermal simulations. *Computational Geosciences*. 4

[5] I. Ben Gharbia (2012). *Résolution de Problèmes de Complémentarité – Application à un Écoulement Diphasique Dans un Milieu Poreux*. PhD Thesis, Université Paris-Dauphine. 10, 12

[6] I. Ben Gharbia, J. Dabaghi, V. Martin, M. Vohralík (2018). A posteriori error estimates and adaptive stopping criteria for a compositional two-phase flow with nonlinear complementarity constraints. Technical report. [hal]. 4

[7] I. Ben Gharbia, E. Flauraud (2019). Study of compositional multiphase flow formulation using complementarity conditions. *Oil & Gas Sciences and Technology*, 74, 1–15. [doi]. 4

[8] I. Ben Gharbia, J.Ch. Gilbert (2012). Nonconvergence of the plain Newton-min algorithm for linear complementarity problems with a P-matrix. *Mathematical Programming*, 134, 349–364. [doi]. 4

[9] I. Ben Gharbia, J.Ch. Gilbert (2013). An algorithmic characterization of P-matricity. *SIAM Journal on Matrix Analysis and Applications*, 34(3), 904–916. [doi]. 4

[10] I. Ben Gharbia, J.Ch. Gilbert (2019). An algorithmic characterization of P-matricity II: adjustments, refinements, and validation. *SIAM Journal on Matrix Analysis and Applications*, 40(2), 800–813. [doi]. 4

[11] I. Ben Gharbia, J. Jaffré (2014). Gas phase appearance and disappearance as a problem with complementarity constraints. *Mathematics and Computers in Simulation*, 99, 28–36. [doi]. 4

[12] J.F. Bonnans (1994). Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29, 161–186. [doi]. 5

[13] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (1997). *Optimisation Numérique – Aspects théoriques et pratiques*. Mathématiques et Applications 27. Springer Verlag, Berlin. 4

[14] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (2006). *Numerical Optimization – Theoretical and Practical Aspects* (second edition). Universitext. Springer Verlag, Berlin. [authors] [editor] [doi]. 4, 11, 12, 21

[15] B. Brogliato (2016). *Nonsmooth Mechanics - Models, Dynamics and Control* (third edition). Springer. [doi]. 4

[16] H. Buchholzer, Ch. Kanzow, P. Knabner, S. Kräutle (2011). The semismooth Newton method for the solution of reactive transport problems including mineral precipitation-dissolution reactions. *Computational Optimization and Applications*, 50(2), 193–221. [doi]. 4

[17] Q.M. Bui, H.C. Elman (2018). Semi-smooth Newton methods for nonlinear complementarity formulation of compositional two-phase flow in porous media. Technical report. [arXiv:1805.05801]. 4

[18] B. Chen, X. Chen, C. Kanzow (2000). A penalized Fischer-Burmeister NCP-function. *Mathematical Programming*, 88, 211–216. [doi]. 4

[19] A. Chiche, J.Ch. Gilbert (2016). How the augmented Lagrangian algorithm can deal with an infeasible convex quadratic optimization problem. *Journal of Convex Analysis*, 23(2), 425–459. [pdf] [editor]. 12

[20] S.J. Chung (1989). NP-completeness of the linear complementarity problem. *Journal of Optimization Theory and Applications*, 60, 393–399. [doi]. 3

[21] F.H. Clarke (1983). *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York. 7

[22] R.W. Cottle (1964). *Nonlinear Programs with Positively Bounded Jacobians*. PhD Thesis, University of California, Berkeley, USA. 4

[23] R.W. Cottle (1966). Nonlinear programs with positively bounded jacobians. *SIAM Journal on Applied Mathematics*, 14, 147–158. [doi]. 4

[24] R.W. Cottle (2005). Linear complementarity since 1978. In *Variational analysis and applications*, Nonconvex Optimization and Its Applications 79, pages 239–257. Springer, New York. [doi]. 3

[25] R.W. Cottle, J.-S. Pang, R.E. Stone (2009). *The Linear Complementarity Problem*. Classics in Applied Mathematics 60. SIAM, Philadelphia, PA, USA. 3

[26] J. Dabaghi, V. Martin, M. Vohralík (2017). Adaptive inexact semismooth Newton methods for the contact problem between two membranes. Technical report, INRIA. [hal-01666845]. 4

[27] T. De Luca, F. Facchinei, C. Kanzow (1996). A semismooth equation approach to the solution of nonlinear complementarity problems. *Mathematical Programming*, 75, 407–439. [doi]. 4

[28] T. De Luca, F. Facchinei, C. Kanzow (2000). A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Computational Optimization and Applications*, 16, 173–205. [doi]. 4, 7

[29] F. Delbos, J.Ch. Gilbert (2005). Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. *Journal of Convex Analysis*, 12(1), 45–69. [preprint] [editor]. 12

[30] V.F. Demyanov, A.M. Rubinov (1983). On quasidifferentiable mappings. *Optimization*, 14(1), 3–21. [doi]. 5

[31] J.E. Dennis, R.B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs. 4, 21

[32] J.-P. Dussault, M. Frappier, J.Ch. Gilbert (2019). A lower bound on the iterative complexity of the Harker and Pang globalization technique of the Newton-min algorithm for solving the linear complementarity problem. *EURO Journal on Computational Optimization*. [doi]. 40

[33] J.-P. Dussault, M. Frappier, J.Ch. Gilbert (2019). A hybrid Newton-min algorithm for the linear complementarity problem (to appear). Technical report. 8, 40

[34] J.-P. Dussault, J.Ch. Gilbert (2019). Asymptotic analysis of a hybrid Newton-min algorithm for complementarity problems (to appear). Technical report. 40

[35] Kenny Erleben (2013). Numerical methods for linear complementarity problems in physics-based animation. In *ACM SIGGRAPH 2013 Courses*, SIGGRAPH '13, pages 8:1–8:42, New York, NY, USA. ACM. URL `http://doi.acm.org/10.1145/2504435.2504443`. 4

[36] F. Facchinei, J.-S. Pang (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems* (two volumes). Springer Series in Operations Research. Springer. 3, 4, 5, 7, 9, 11

[37] M.C. Ferris, J.-S. Kanzow (2002). Complementarity and related problems. In P.M. Pardalos, M.G.C. Resende (editors), *Handbook of Applied Optimization*. Oxford University Press, New York, NY, USA. 4

[38] M.C. Ferris, J.-S. Pang (1997). Engineering and economic applications of complementarity problems. *SIAM Review*, 39, 669–713. [doi]. 4

[39] A. Fischer, J. Houyuan (2000). Merit functions for complementarity and related problems: a survey. *Computational Optimization and Applications*, 17, 159–182. [doi]. 4

[40] Mathieu Frappier. Reformulation semi-lisse appliquée au problème de complémentarité. Master's thesis, Département de Math'ematiques, Facul'e des sciences, Université de Sherbrooke, (2019, January). 8

[41] S.A. Gabriel, J.-S. Pang (1993). NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60, 295–337. [doi]. 9, 10

[42] A. Galántai (2012). Properties and construction of NCP functions. *Computational Optimization and Applications*, 52(3), 805–824. 4

[43] Y. Gao, H. Song, X. Wang, K. Zhang (2020). Primal-dual active set method for pricing American better-of option on two assets. *Communications in Nonlinear Science and Numerical Simulation*, 80. [doi]. 4

[44] J.Y. Han, D.F. Sun (1997). Newton and quasi-Newton methods for normal maps with polyhedral sets. *Journal of Optimization Theory and Applications*, 94, 659–676. [doi]. 7

[45] S.-H. Han, J.-S. Pang, N. Rangaraj (1992). Globally convergent Newton methods for nonsmooth equations. *Mathematics of Operations Research*, 17, 586–607. [doi]. 6, 35

[46] P.T. Harker, J.-S. Pang (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48, 161–220. [doi]. 4

[47] T. Hoheisel, Ch. Kanzow, B.S. Mordukhovich, H. Phan (2012). Generalized Newton's method based on graphical derivatives. *Nonlinear Analysis*, 75(3), 1324–1340. [doi]. 4

[48] G. Isac (1992). *Complementarity Problems.* Lecture Notes in Mathematics 1528. Springer, Berlin. [doi]. 3, 4

[49] K. Ito, K. Kunisch (2009). On a semi-smooth Newton method and its globalization. *Mathematical Programming*, 118, 347–370. [doi].

[50] A.F. Izmailov, M.V. Solodov (2014). *Newton-Type Methods for Optimization and Variational Problems.* Springer Series in Operations Research and Financial Engineering. Springer. 3, 4, 7

[51] N.H. Josephy (1979). Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI, USA. 5

[52] C. Kanzow, H. Kleinmichel (1998). A new class of semismooth Newton-type methods for nonlinear complementarity problems. *Computational Optimization and Applications*, 3, 227–251. [doi]. 4

[53] C. Kanzow, N. Yamashita, M. Fukushima (1997). New NCP-functions and their properties. *Journal of Optimization Theory and Applications*, 94(1), 115–135. [doi]. 4

[54] S. Karamardian (1971). Generalized complementarity problem. *Journal of Optimization Theory and Applications*, 8(3), 161–168. [doi]. 5

[55] M. Kojima, N. Megiddo, T. Noma, A. Yoshise (1991). *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems.* Lecture Notes in Computer Science 538. Springer, Berlin. [doi]. 3

[56] M. Kojima, S. Shindo (1986). Extension of Newton and quasi-Newton methods to systems of PC^1 equations. *Journal of Operations Research Society of Japan*, 29, 352–375. 7, 10, 16

[57] S. Kräutle (2011). The semismooth Newton method for multicomponent reactive transport with minerals. *Advances in Water Resources*, 34(1), 137–151. [doi]. 4

[58] B. Kummer (1988). Newton's method for nondifferentiable functions. In J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach, K. Zimmerman (editors), *Advances in Mathematical Optimization*, pages 114–125. Akademie-Verlag, Berlin. 10, 16

[59] X. Li, D. Sun, K.-C. Toh (2018). On the efficient computation of a generalized Jacobian of the projector over the Birkhoff polytope. Technical report. arXiv:1702.05934. 7

[60] V.L.R. Lopes, J.M. Martínez, R. Pérez (1999). On the local convergence of quasi-newton methods for nonlinear complementarity problems. *Applied Numerical Mathematics*, 30, 3–22. 7

[61] M.-A. Majig, M. Fukushima (2010). Restricted-step josephy-newton method for general variational inequalities with polyhedral constraints. *Pacific Journal of Optimization*, 6(2), 375–390. 5

[62] O.L. Mangasarian (1976). Equivalence of the complementarity to a system of nonlinear equations. *SIAM Journal on Applied Mathematics*, 31(1), 89–92. [doi]. 4

[63] O.L. Mangasarian, S. Fromovitz (1967). The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17, 37–47. [doi]. 22

[64] E. Marchand, T. Müller, P. Knabner (2012). Fully coupled generalised hybrid-mixed finite element approximation of two-phase two-component flow in porous media. Part II: numerical scheme and numerical results. *Computational Geosciences*, 16(3), 691–708. [doi]. 4

[65] E. Marchand, T. Müller, P. Knabner (2013). Fully coupled generalised hybrid-mixed finite element approximation of two-phase two-component flow in porous media. Part I: formulation and properties of the mathematical model. *Computational Geosciences*, 17(2), 431–442. [doi]. 4

[66] K.G. Murty (1988). *Linear Complementarity, Linear and Nonlinear Programming* (Internet edition, prepared by Feng-Tien Yu, 1997). Heldermann Verlag, Berlin. 3

[67] J.M. Ortega, W.C. Rheinboldt (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York. 4

[68] J.-S. Pang (1990). Newton's method for B-differentiable equations. *Mathematics of Operations Research*, 15, 311–341. [doi]. 5, 6

[69] J.-S. Pang (1991). A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Mathematical Programming*, 51(1-3), 101–131. [doi]. 4, 6, 7, 16, 35

[70] J.-S. Pang (1995). Complementarity problems. In R. Horst, P.M. Pardalos (editors), *Handbook of Global Optimization*, volume 2 of *Nonconvex Optimization and Its Applications*, pages 271–338. Kluwer, Dordrecht. [doi]. 3, 4

[71] S. Pieraccini, M.G. Gasparo, A. Pasquali (2003). Global Newton-type methods and semismooth reformulations for NCP. *Applied Numerical Mathematics*, 44, 367–384. [doi]. 4, 7

[72] L. Qi (1993). Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18, 227–244. 7

[73] L. Qi, J. Sun (1993). A nonsmooth version of Newton's method. *Mathematical Programming*, 58, 353–367. [doi]. 7

[74] S.M. Robinson (1980). Strongly regular generalized equations. *Mathematics of Operations Research*, 5, 43–62. [doi]. 5

[75] S.M. Robinson (1987). Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity. *Mathematical Programming Study*, 30, 45–66. [doi]. 5

[76] A. Shapiro (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66, 477–487. 5

[77] D. Sun, L. Qi (1999). On NCP-functions. *Computational Optimization and Applications*, 13(1-3), 201–220. [doi]. 4

[78] K. Ueda, N. Yamashita (2012). Global complexity bound analysis of the Levenberg-Marquardt method for nonsmooth equations and its application to the nonlinear complementarity problem. *Journal of Optimization Theory and Applications*, 152, 450–467. 40